

**Universidad de Buenos Aires**



**Facultad de Ciencias Exactas y Naturales**

**Maestría en Explotación de Datos y Descubrimiento del  
Conocimiento**



**Plan de Tesis de Maestría**

**Pronósticos de precipitación semanal probabilística  
calibrados sobre el Sur de Sudamérica**

**Esp. Alfredo Luis Rolla**

**Director de Tesis (Maestría): Mg. Ing. Gustavo Denicolay**

**Co-directora de Tesis: Dra. Carolina Susana Vera**

## **Tema de investigación.**

Diversos sectores socio-económicos son afectados por fenómenos meteorológico-climáticos con periodicidades de varias semanas para los cuales no existen en el país productos operacionales. Sin embargo, desde hace algunos años los modelos globales como los del NCEP de los Estados Unidos proporcionan pronósticos dinámicos de frecuencia semanal que podrían ser utilizados a escala regional si se les aplica los tratamientos matemáticos adecuados. El desarrollo propuesto utilizará modelos de pronósticos dinámicos como CFSv2 (NCEP Climate Forecast System version 2) (Saha et al., 2014) y GEFS (NOAA Global Ensemble Forecast System) (Hamill et al., 2013) y aplicará métodos de aprendizaje automático de calibración supervisados para eliminar los errores sistemáticos de los modelos dinámicos. El carácter probabilístico y calibrado del producto, sumado a una verificación del desempeño en la región de interés, contribuirá a elevar el valor socio-económico de los mismos para la toma de decisión.

En consecuencia, se plantean los siguientes objetivos específicos:

1. Evaluar distintas bases de datos de distribuciones espaciales históricas de observaciones de precipitación semanal en la región.
2. Evaluar la calidad de los distintos modelos de calibración para diferentes umbrales de precipitación aplicados a los modelos de pronósticos semanales mediante índices de verificación adecuados.
3. Realizar pruebas de estabilidad de los parámetros de estos modelos de calibración respecto del tamaño espacial de la muestra.
4. Desarrollar una herramienta de visualización en tiempo operacional de tipo web de mapas probabilísticos de precipitación semanal para distintos umbrales en un horizonte temporal de la semana 2 a la semana 4, como ayuda para la toma de decisión.

## **Antecedentes sobre el tema.**

De manera simplista, los usuarios de los pronósticos meteorológicos toman o no toman decisiones dependiendo de si se pronostica o no un evento meteorológico en particular, por ejemplo, introducen acciones de protección para prevenir/reducir las pérdidas relacionadas con el clima.

Se podría aplicar un análisis de ganancia-pérdida de diferente complejidad para evaluar el impacto económico del uso de los pronósticos meteorológicos por los usuarios (Murphy 1985; Katz y Murphy 1997). Los estudios sobre el valor económico de los pronósticos meteorológicos pueden ser descriptivos, evaluando el valor de los pronósticos utilizados, a menudo de manera subóptima, por los usuarios; o prescriptivo, identificando el valor potencial de los pronósticos, asumiendo que se usan de manera óptima (Stewart 1997).

La inclusión de la incertidumbre en los pronósticos meteorológicos y climáticos puede llevar a un aumento sustancial de su valor económico (Zhuet al., 2002; Palmer et al., 2007), en temas de hidrología, agricultura, industria, entre otros, así como mejorar la información necesaria para la gestión de emergencias producidas por fenómenos hidrometeorológicos.

Tanto los pronósticos meteorológicos (de frecuencia diaria) como los climáticos (de frecuencia semanal, trimestral e inclusive de años) se realizan utilizando modelos matemáticos de dimensión global que resuelven las ecuaciones fundamentales que rigen los procesos físicos

que se producen en la atmósfera, océanos, hielos, suelo, etc. Estos modelos se los denomina generalmente modelos globales y, “pronósticos dinámicos” a las salidas que proporcionan cuando sus condiciones iniciales se definen a partir de las observaciones meteorológicas.

Durante las últimas décadas, el aumento de los recursos informáticos de cómputo como clusters de computadoras y procesamiento paralelo llevó al desarrollo de modelos dinámicos atmosféricos más realistas y el reconocimiento de la importancia de la predictibilidad atmosférica en general, debido a esto, los pronósticos por ensambles se convirtieron en un componente importante de la Predicción Numérica del Tiempo.

¿Por qué el enfoque de ensambles es exitoso? un ensamble de pronósticos naturalmente ofrece una multitud de niveles de decisión en comparación con una sola decisión de sí/no basada en un solo pronóstico de control, proporcionando distribuciones de probabilidad detalladas en lugar de solo dos niveles de probabilidad.

Como mostró Toth et al. (1998), el uso de distribuciones de probabilidad detalladas basadas en ensambles (en comparación con el uso de solo dos niveles de probabilidad) mejora sustancialmente el desempeño del pronóstico en términos de ROC, y otros indicadores de desempeño. Para mejorar los pronósticos dinámicos por ensambles de variables meteorológicas estos deben ser post-procesados estadísticamente para eliminar posibles errores sistemáticos o sesgos del modelo antes de que se usen en el pronóstico del tiempo. La calibración es un tema importante para las aplicaciones prácticas ya que, como mostró Wilks (2001), los pronósticos no calibrados pueden sufrir una gran reducción en su valor socio-económico esperado.

Para corregir el efecto de los errores sistemáticos del ensamble, en el caso de pronósticos de frecuencia diaria se han desarrollado varias técnicas, todas ellas basadas en el estudio de la relación entre el error y el valor de pronóstico y en el desarrollo de modelos estadísticos para calcular una probabilidad calibrada dados los pronósticos de miembros del ensamble (Ruiz et al. 2012, Hamill and Colucci, 1997, 1998; Eckel and Walters, 1998; Applequist et al., 2002; Gahrs et al., 2003; Gallus and Seagal, 2004; Raftery et al., 2005; Hamill and Whitaker, 2006; McLean Slaughter et al., 2007; Stensrud and Yussouf, 2007, entre otros). Estas técnicas no fueron aplicadas a pronósticos de frecuencia semanal en la región de interés, que es el objetivo de este proyecto.

### **Aporte esperado al finalizar el proyecto:**

Es fundamental que los usuarios tengan acceso a información probabilística que capture las variaciones en la incertidumbre esperada de los pronósticos semanales. De hecho, un pronóstico del tiempo no está completo a menos que se exprese en forma de distribuciones o mapas de probabilidad de ocurrencia de precipitación, en nuestro caso, semanal. En cuanto a la incertidumbre, el objetivo de la predicción semanal incluido el posprocesamiento estadístico, será la provisión de una distribución de probabilidad detallada dependiente del caso, y no sólo la mejor estimación del estado de la atmósfera por una sola corrida de control. Dicha información adicional facilitaría el uso y aumentaría el valor socio-económico potencial de los pronósticos meteorológicos semanales en la región.

Un gran desafío de este trabajo es la comunicación de estos pronósticos probabilísticos mediante visualizaciones usando interfaces simples destinadas al usuario final quien dará el mejor uso posible de los mismos.

De este modo los usuarios de estos pronósticos semanales calibrados de precipitación , podrán tomar esta información, junto con otros factores, e incorporarlos a sus decisiones relacionadas con sus acciones sensibles al clima.

### **Transferencias de los resultados.**

Los resultados de este trabajo podrían ser utilizados/transferidos al Servicio Meteorológico Nacional, para aplicarlo como una herramienta más dentro de sus servicios. Y al ser un producto operativo podría ser utilizado por aquellos usuarios interesados en mejorar sus decisiones usando pronósticos que manejen las incertezas de los pronósticos.

### **Lugar de trabajo**

El trabajo se desarrollará en:

Centro de Investigaciones del Mar y la Atmósfera. CIMA / CONICET-UBA  
Intendente Güiraldes 2160 - Ciudad Universitaria - Pabellón II - 2do. piso  
(C1428EGA) Buenos Aires - Argentina  
TE: (54)(11) 5285-8467 - Fax: (54)(11) 4788-3572  
Email: [webmaster@cima.fcen.uba.ar](mailto:webmaster@cima.fcen.uba.ar)

### **Disponibilidad de infraestructura, factibilidad de desarrollo del trabajo**

- Un Cluster con 368 procesadores con MPICH2 1.4.1p1 (libs de paralelización ), Torque 3.0.3 (colas) y Conectividad: 1000 Mbps.
- Acceso a una placa de tipo GPU NVIDIA 1070 ti con 2400 cores y 8GB de memoria.
- Un servidor dedicado con 12 procesadores y 16GB de memoria para el procesamiento de la información de los modelos CFSv2 y GEFS con 4TB de almacenamiento HDD y un disco de 512GB de tipo SSD.
- Sistema de almacenamiento masivo de datos (data storage) con una capacidad total de 300 TB.
- Un servidor web para el desarrollo de la aplicación propuesta.
- El servicio de internet de alta velocidad es provisto por la Facultad de Ciencias Exactas y Naturales (UBA) y con acceso a Internet II (Redes Avanzadas, InnovaRed).

### **Plan de trabajo y cronograma tentativo.**

Para realizar la calibración de los pronósticos dinámicos es necesario contar con distribuciones espaciales históricas de observaciones de precipitación como así también con corridas de los modelos de pronóstico dinámicos históricas (también denominadas 'hindcast').

En el caso de los pronósticos dinámicos existen 'hindcast' disponibles para los modelos considerados CFSv2 y GEFS. El tema de las observaciones es algo a evaluar, ya que si bien existen observaciones en puntos de estaciones meteorológicas con registros largos pertenecientes al Servicio Meteorológico Nacional (SMN) o del Instituto Nacional de Tecnología Agropecuaria (INTA). La distribución espacial no es muy buena para realizar la calibración de los modelos dinámicos. Por lo tanto durante el trabajo se evaluarán no sólo datos en estaciones meteorológicas con registros largos sino información de estimación de precipitación de tipo satelital. Asimismo, hace unas décadas surgieron diversas técnicas para estimar la precipitación acumulada en base a sensores remotos y a una combinación de sensores remotos y pluviómetros (Joyce et al., 2004; Huffman et al., 2007; entre otros) que buscan lograr una descripción detallada de la variabilidad temporal y espacial de la precipitación aún en aquellas regiones donde la densidad de estaciones es muy escasa. Uno de los mayores avances en este campo, fue la incorporación en varios satélites (TRMM, SSM/I, NOAA, IMERGE) de sensores en microondas pasivas que proveen mejores estimaciones de la tasa de precipitación (Kidd et al., 2003). Se evaluará el uso de datasets de precipitación grilladas de estaciones meteorológicas y estimadas por satélite con TRMM (Huffman et al., 2007) y CMORPH (CPC Morphing Technique) (Joyce et al., 2004) y otros. Estas estimaciones y otras están disponibles en internet y comprenden el periodo de los 'hindcast' de los modelos dinámicos de pronóstico.

En el caso de las estimaciones de satélite ('observaciones') y los modelos de pronóstico dinámico será necesario construir datasets semanales para la aplicación de las calibraciones a los ensambles de frecuencia semanales.

Se evaluarán/construirán modelos de calibración para distintos umbrales de precipitación usando regresión logística (Cox, 1958) usando los miembros del ensamble, regresión logística extendida (Wilks, 2009) usando además la variabilidad del ensamble, calibración beta (Kull M. et al., 2017) agregando un parámetro más al modelo de regresión logística variando la forma de la misma, histogramas de rangos (Hamill and Colucci, 1998) y otros.

Se realizarán pruebas de sensibilidad de estabilidad de los parámetros de estos modelos respecto del tamaño espacial de la muestra.

Se realizará la verificación de las calibraciones usando índices como BSS ('brier skill score') (Brier, 1950) y gráficos de diagnóstico como los 'reliability diagrams' (Hartmann et al. 2002) para evaluar los pronósticos probabilísticos resultantes.

Resulta evidente que este tipo de pronósticos semanales no son conocidos ni por los pronosticadores del Servicio Meteorológico Nacional por ejemplo, ni por los usuarios. Es por eso que resulta necesario desarrollar herramientas para su visualización que luego se utilizarán en modo experimental como se ha realizado previamente con los productos que se disponibilizan en el portal [climar.cima.fcen.uba.ar](http://climar.cima.fcen.uba.ar). Se desarrollará una herramienta/aplicación de tipo web para la visualización de los pronósticos probabilísticos semanales en distintos umbrales y despliegue de mapas de pronósticos probabilísticos de precipitación. Para la implementación de la misma se usarán en principio herramientas de tipo open-source, como el servidor web Apache (The Apache Software Foundation. 2019), usando programación javascript y todos los frameworks disponibles aplicables a solucionar este problema para manejo de información georreferenciada (openLayers) y para el diseño de la interface (jquery y otros). Para el back-end se usará una base de datos open-source como MySQL con lenguaje PHP.

## **Actividades específicas:**

### *A. Armado de los datasets semanales de observaciones y pronósticos en la región:*

- (1) Descarga de los datasets de los ensambles de los modelos CFSv2 y GEFS.
- (2) Descarga de los dataset de observaciones CPC-UNI, TRMM (3B42RT) y CMORPH y otros.
- (3) Transformar los dataset horarios originales en archivos de precipitación semanales.

### *B. Armado de los modelos de calibración:*

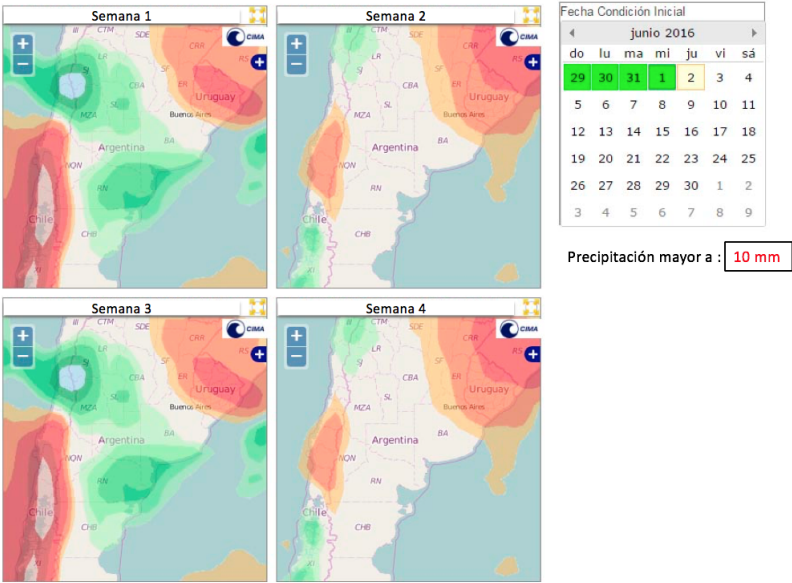
- (1) Implementar los modelos calibración de Regresión logística, Regresión logística extendida, distribución Beta y otros.
- (2) Probarlos en un par de localidades testigos y evaluar la regionalización de N puntos para la construcción de los modelos de calibración.
- (3) Determinar los umbrales de precipitación a utilizar para los modelos.
- (4) Verificar los modelos en las localidades seleccionadas como testigos, usando curvas ROC, AUC y Diagramas de confiabilidad así como 'Brier Score' en base a los umbrales de precipitación y regionalización seleccionados.
- (5) Optimizar los parámetros de los modelos de calibración en base a la cantidad de datos incorporados en la regionalización para su construcción.
- (6) Generalizar a toda la región los modelos que mejor se comportaron en base a los índices de verificación usados.

### *C. Implementar en tiempo operacional los mapas de pronósticos probabilísticos semanales:*

- (1) Diseño e implementación de la interface web para el despliegue de los mapas de pronósticos probabilísticos (La interface incluirá la selección del umbral, el modelo (CFSv2 o GEFS) y desplegará el mapa de probabilidades de precipitación.
- (2) Implementación del script de descarga diario de los pronósticos operativos CFSv2 y GEFS.
- (3) Cómputo de los mapas probabilísticos usando los modelos de calibración de precipitación en los distintos umbrales para la región.
- (4) Creación de los mapas de probabilidades georreferenciados para ser mostrados en la interface web.

Maqueta de interface web , para los pronósticos semanales :

Pronostico probabilístico semanal de precipitación usando un ensamble medio de 16 miembros del modelo NOAA\_NCEP-CFSv2 ( Referencia 01-07-2019)



Cronograma de Actividades:

Cronograma tentativo de Actividades						
Actividad	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6
A.1	■					
A.2	■					
A.3	■					
B.1		■	■			
B.2		■	■			
B.3		■	■			
B.4		■	■			
B.5				■		
B.6				■		
C.1					■	
C.2					■	
C.3						■
C.4						■

## Referencias bibliográficas

Applequist S, Gahrs EG, Pfeffer RL, Niu X-F. 2002. Comparison of methodologies for probabilistic quantitative precipitation forecasting. *Weather and Forecasting* 17: 783–799.

Brier GW. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78: 1–3.

Cox, DR (1958). "The regression analysis of binary sequences (with discussion)". *J Roy Stat Soc B*. 20 (2): 215–242.

Eckel FA, Walters MK. 1998. Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. *Weather and Forecasting* 13: 1132–1147.

Gahrs GE, Applequist S, Pfeffer RL, Niu X. 2003. Improved results for probabilistic quantitative precipitation forecasting. *Weather and Forecasting* 18: 879–890.

Gallus WA, Seagal M. 2004. Does increased predicted warm-season rainfall indicate enhanced likelihood of rain occurrence? *Weather and Forecasting* 19: 1127–1135.

Hamill T, Colucci SJ. 1997. Verification of Eta-RSM short-range ensemble forecasts. *Monthly Weather Review* 125: 1312–1327.

Hamill T, Colucci SJ. 1998. Evaluation of Eta-RSME ensemble probabilistic precipitation forecasts. *Monthly Weather Review* 126: 711 – 724.

Hamill T, Whitaker JS. 2006. Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Monthly Weather Review* 134: 3209–3229.

Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau Jr., Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, 94, 1553–1565.

Hartmann, H.C., Pagano, T.C., Sorooshian, S. and Bales, R. (2002). Confidence builder: evaluating seasonal climate forecasts from user perspectives. *Bull Amer. Met. Soc.*, 84, 683-698.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., ... Stocker, E. F. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *Journal of Hydrometeorology*, 8(1), 38–55. doi:10.1175/jhm560.1

Joyce RJ, Janowiak JE, Arkin PA, Xie P. 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology* 5: 487–503.

Katz, R. W., and A. H. Murphy, Eds., 1997: *Economic Value of Weather and Climate Forecasts*. Cambridge University Press, 222 pp.

Kull, M., De Menezes E Silva Filho, T., & Flach, P. (2017). Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017) (JMLR Workshop and Conference Proceedings; Vol. 54)*. *Journal of Machine Learning Research*.



McLean Sloughter J, Raftery A, Gneiting T, Fraley C. 2007. Probabilistic quantitative precipitation forecasting using bayesian model averaging. *Monthly Weather Review* 135: 3209–3220.

Murphy, A. H., 1985: Decision making and the value of forecasts in a generalized model of the cost–loss ratio situation. *Mon. Wea. Rev.*, **113**, 362–369.

Palmer TN, Buizza R, Leutbecher M, Hagedorn R, Jung T, Rodwell M, Vitart F, Berner J, Hagel E, Lawrence A, Pappenberger F, Park Y-Y, von Bremen L, Gilmour I. 2007. The Ensemble Prediction System – Recent and Ongoing Developments. ECMWF Technical Memoranda N°430, October 2007, 53 pp.

Raftery A, Gneiting T, Balabdaoui F, Polakowski M. 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* 133: 1155–1174.

Ruiz J. J., Saulo C., 2012. How sensitive are probabilistic precipitation forecasts to the choice of calibration algorithms and the ensemble generation method? Part I: sensitivity to calibration methods. *Meteor. Appl.*, 19: 302–313 .

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., ... Becker, E. (2014). The NCEP Climate Forecast System Version 2. *Journal of Climate*, 27(6), 2185–2208. doi:10.1175/jcli-d-12-00823.1

Stensrud D, Yussouf N. 2007. Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecasting system. *Weather and Forecasting* 22: 3–17.

Stewart, T. R., 1997: Forecast value: Descriptive decision studies. *Economic Value of Weather and Climate Forecasts*, R. W. Katz and A. H. Murphy, Eds., Cambridge University Press, 147–181.

Toth, Z., Y. Zhu, T. Marchok, S. Tracton, and E. Kalnay, 1998: Verification of the NCEP global ensemble forecasts. Preprints, *12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, Amer. Meteor. Soc., 286–289.

Wilks, D. S., 2001: A skill score based on economic value for probability forecasts. *Meteor. Appl.*, **8**, 209–219.

Wilks, D. S. 2009. Extending logistic regression to provide full probability distribution MOS forecasts. *Meteorological Applications* 16: 361–368.

Zhu Y, Toth Z, Wobus R, Richardson D, Mylne K. (2002). The economic value of ensemble-based weather forecast. *Bulletin of the American Meteorological Society* 83: 73–83.