



Pronóstico de precipitación mensual en Argentina aplicando técnicas de machine learning

Dra. Garbarini, E.; Dra. González, M.; Mg. Rolla, A.

- Introducción
- Esquema del pronóstico
- Conclusiones

Introducción

Pronóstico a mediano plazo

- Acorta las distancias entre el pronóstico del tiempo y del clima.
- Tiene grandes dificultades en términos de habilidad de pronóstico.
- El desafío proviene de las diferentes escalas temporales y espaciales de donde proviene la predictibilidad y la capacidad de predecirlas correctamente.

PRONÓSTICOS DEL TIEMPO

La predictibilidad proviene de las condiciones iniciales atmosféricas

PREDICCIONES S2S

La predictibilidad proviene de las condiciones iniciales atmosféricas, condiciones del suelo, mar, hielo, la estratósfera y otras fuentes.

PERSPECTIVA ESTACIONAL

La predictibilidad proviene principalmente de las condiciones de la TSM, como por ejemplo ENSO

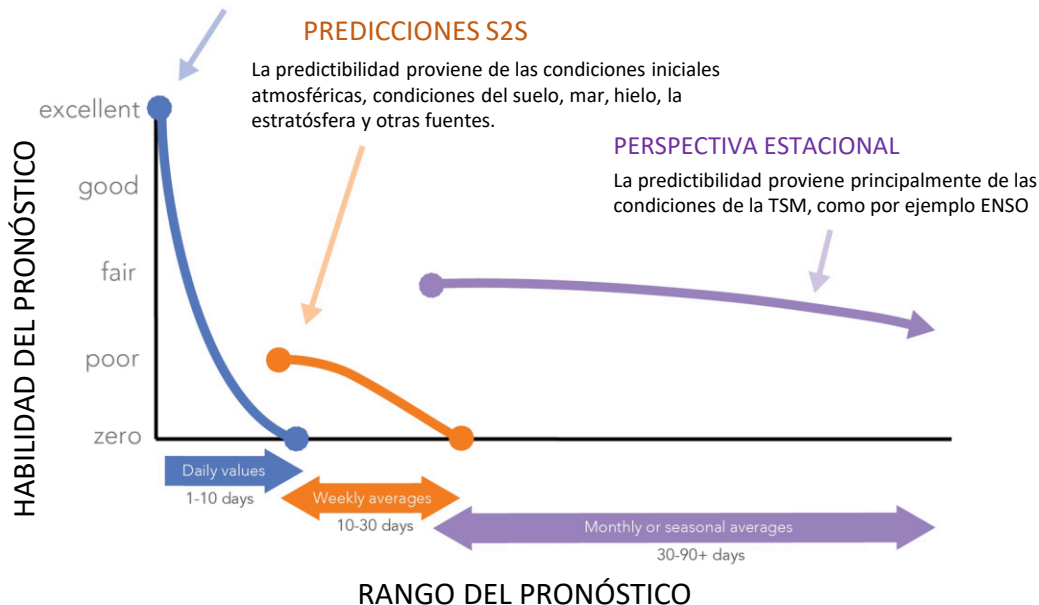


Fig. 1: Esquema de la habilidad del pronóstico en función de las escalas de tiempo (adaptado de White et al., 2017)

Etapas del desarrollo del modelo de pronóstico

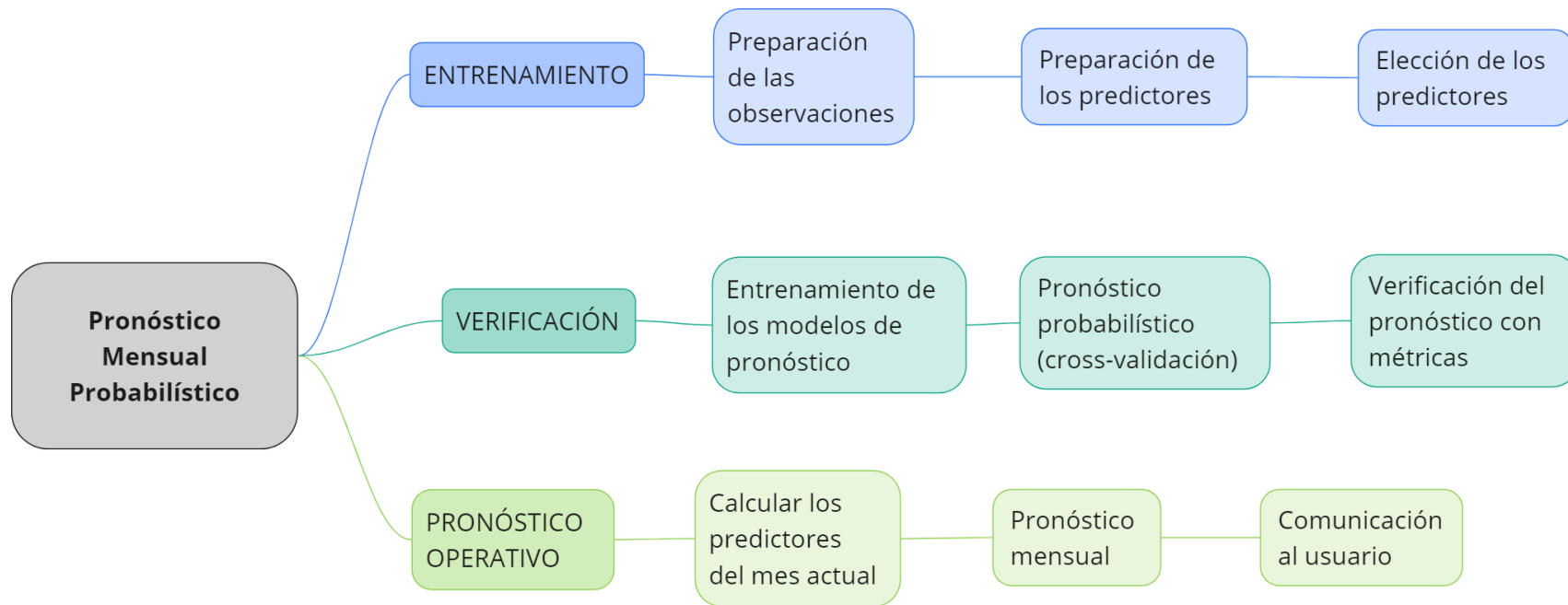


Fig. 2: Esquema del desarrollo del pronóstico y pasos de trabajo.

1 Entrenamiento

1. Preparación de las observaciones

- 91 estaciones meteorológicas de Argentina (SMN, INTA) para 1979-2023.
- Se delimitaron 6 regiones a partir del cálculo de ondas anuales.
- Se crearon series medias de precipitación mensual como el promedio espacial de las estaciones meteorológicas que componen cada región.

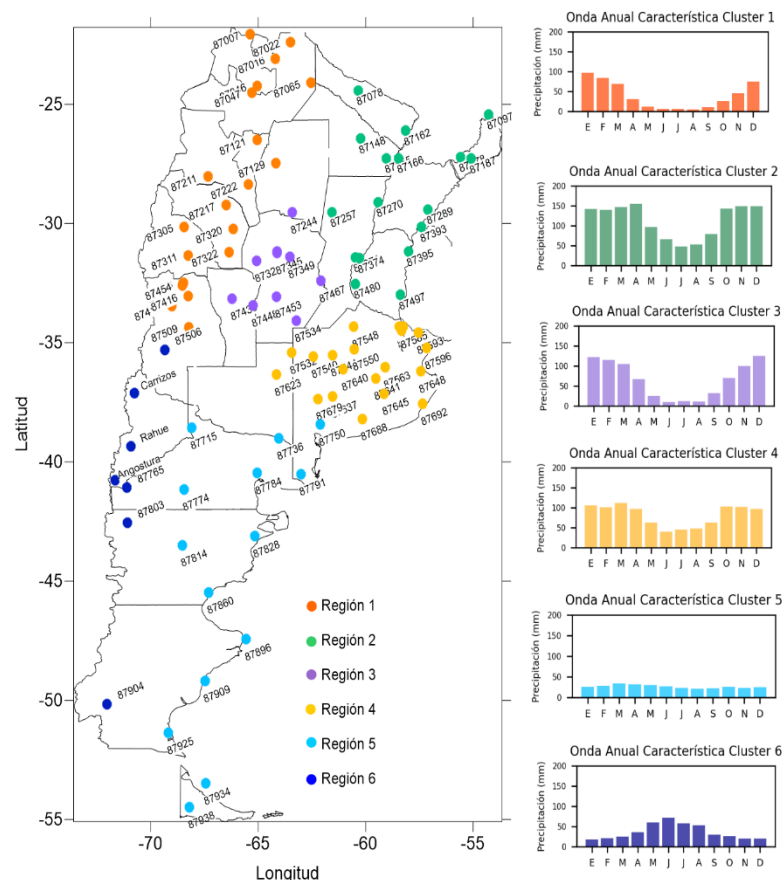


Fig.3: Estaciones regionalizadas junto a su onda anual característica.

1 Entrenamiento

2. Preparación del reanálisis y elección de los predictores

- Datos mensuales de NCEP/NCAR desde 1981 en adelante
(Resolución espacial: retículas de 250 x 250 km²)

- HGT200: Altura geopotencial en 200 hPa
- HGT500: Altura geopotencial en 500 hPa
- HGT1000: Altura geopotencial en 1000 hPa
- SST: Temperatura superficie del mar
- TCW: Agua total en la columna
- U850: Viento zonal en 850 hPa
- V850: Viento meridional en 850 hPa

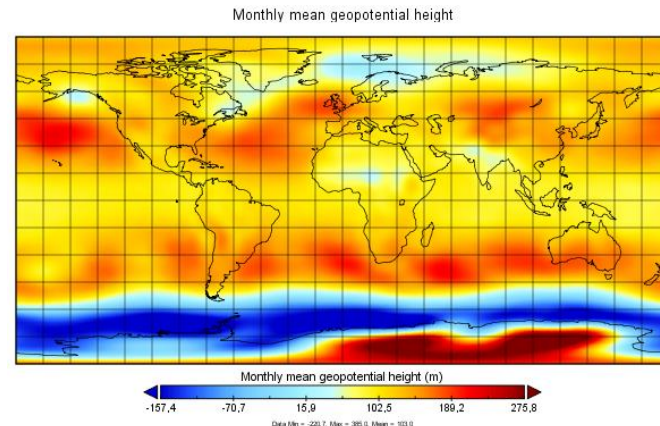
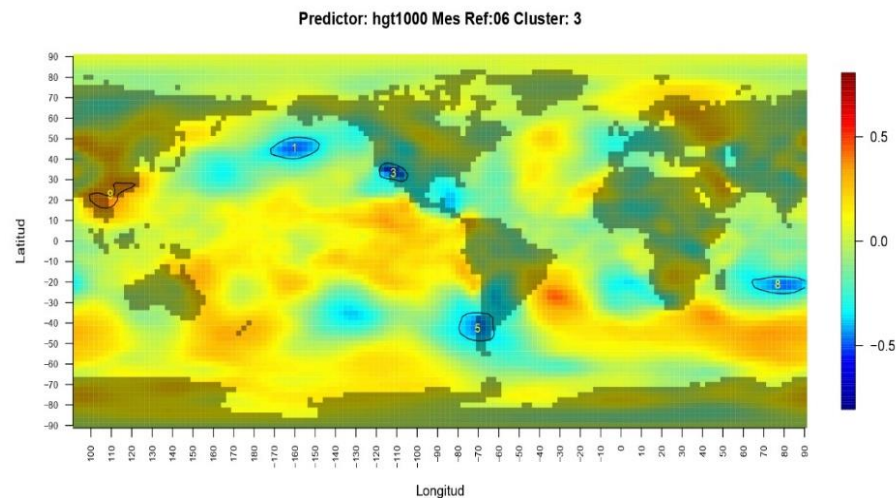


Fig. 4: Campo de altura geopotencial en 1000 hPa para un dado mes y año.

1 Entrenamiento

2. Preparación del reanálisis y elección de los predictores

- Se correlacionaron las series de precipitación de cada región y mes del año con las series de los predictores del mes previo con un 95% de confianza (Fig. 5)
- Se filtra manualmente aquellos que carezcan de sentido físico.
- LASSO: reducir la cantidad de predictores significativos y obtener el mejor conjunto de predictores independientes y físicamente consistentes.

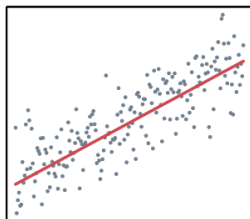


2 Verificación

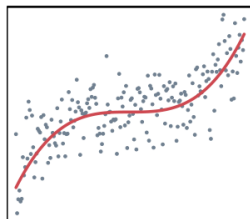
1. Entrenamiento de los modelos

- Se construyeron los modelos teniendo en cuenta las metodologías:

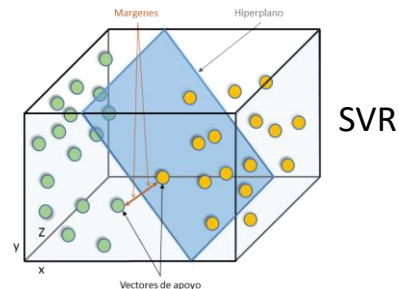
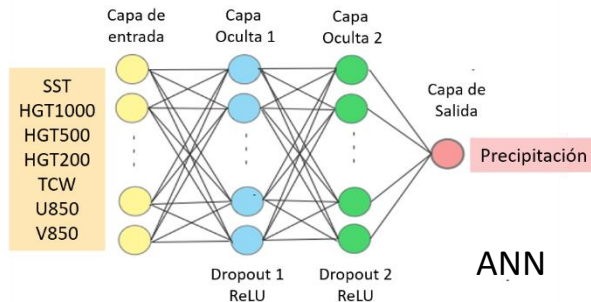
- RLM: Linear Regression Models
- GAM: General Additive Models
- ANN: Artificial Neural Networks
- SVR: Support Vector Machines
- ENS: Ensamble de las 4 metodologías



RLM



GAM



2. Pronóstico Probabilístico y cross-validación

- Para cada mes y cluster: el programa lee los predictores seleccionados y los modelos obtenidos en las corridas anteriores y devuelve el pronóstico, considerando todos los modelos que expliquen al menos el 50% de la precipitación.
- El periodo de entrenamiento inicial fue 1981-2015, y los modelos se entrenaron iterativamente avanzando un año en el pronóstico.

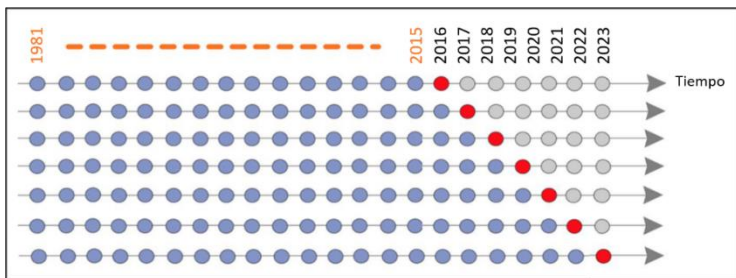


Fig. 7: Esquema del pronóstico y los períodos de entrenamiento

A	B	C	D	E	F	G	H
N.Modelo	adjR	Umbral	Pred.Año	Formula	Tipo	Prono	obs
1	0,55219924	0,5	2023	y.dato ~ sst_	RLM	33,73	22,0568571
2	0,56792074	0,5	2023	y.dato ~ sst_	RLM	26,51	22,0568571
3	0,57606956	0,5	2023	y.dato ~ sst_	RLM	34,11	22,0568571
4	0,5414356	0,5	2023	y.dato ~ sst_	RLM	32,6	22,0568571
5	0,57946648	0,5	2023	y.dato ~ sst_	RLM	28,6	22,0568571
6	0,55630599	0,5	2023	y.dato ~ sst_	RLM	26	22,0568571
7	0,56492666	0,5	2023	y.dato ~ sst_	RLM	33,33	22,0568571
8	0,56765431	0,5	2023	y.dato ~ sst_	RLM	28,22	22,0568571
9	0,656	0,5	2023	y.dato ~ s(sst	GAM	27,05	22,0568571
10	0,653	0,5	2023	y.dato ~ s(sst	GAM	22,05	22,0568571
11	0,652	0,5	2023	y.dato ~ s(sst	GAM	27,8	22,0568571
12	0,65	0,5	2023	y.dato ~ s(sst	GAM	25,99	22,0568571
13	0,648	0,5	2023	y.dato ~ s(sst	GAM	24,29	22,0568571
14	0,648	0,5	2023	y.dato ~ s(sst	GAM	20,46	22,0568571

J	K	L	M	N	O	P	Q	R
Quintiles	P.mayor.que	lim.inf	Proba		Quintiles	P.entre	Proba	obs
6.03 - 15.73	P[pre > 6.03]	6,02857143	0,99		6.03 - 15.73	P[pre > 6.03]	0,08	22,0568571
15.73 - 22.20	P[pre > 15.73]	15,7257143	0,91		15.73 - 22.20	P[pre > 15.73]	0,17	
22.20 - 28.34	P[pre > 22.20]	22,2028571	0,74		22.20 - 28.34	P[pre > 22.20]	0,34	
28.34 - 39.21	P[pre > 28.34]	28,3442857	0,4		28.34 - 39.21	P[pre > 28.34]	0,34	
39.21 - 73.61	P[pre > 39.21]	39,2102857	0,06		39.21 - 73.61	P[pre > 39.21]	0,06	
Tipo	N	MEDIA	DESvío					
ANN	4	25,4825	2,18754619					
GAM	28	18,9267857	5,81767807					
RLM	8	30,3875	3,39666118					
SVR	48	29,4410417	5,46932889					
TOTAL	88	26,0017045	7,22317379					

Fig. 6: Salida del programa de Pronóstico probabilístico para el mes de marzo de 2023 y el Cluster 6 (Patagonia Andina).

2 Verificación

3. Métricas de performance – Gráficos de dispersión

- Buena habilidad en general en marzo, julio, agosto, septiembre y noviembre.
- Patagonia E (C5): buena performance todo el año.
- Patagonia O (C6): buena performance cuando la precipitación es escasa (septiembre a abril)
- Semestre cálido: mayor dispersión en
 - NOA (C1) en enero y febrero
 - Litoral (C2) de octubre a diciembre

- C1: NOA
- C2: NEA
- C3: Centro del país
- C4: Buenos Aires
- C5: Patagonia E
- C6: Patagonia O

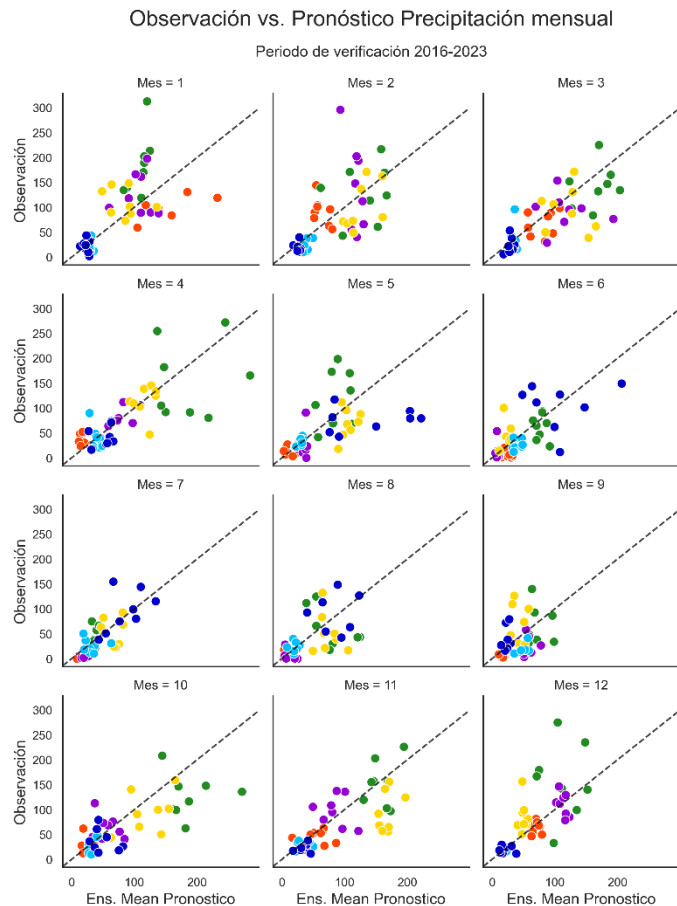


Fig. 8: Diagrama de dispersión del valor observado versus valor pronosticado junto con la línea de identidad (línea punteada) utilizando el ensamble para cada mes del periodo de verificación.

2 Verificación

3. Gráficos de dispersión por metodología

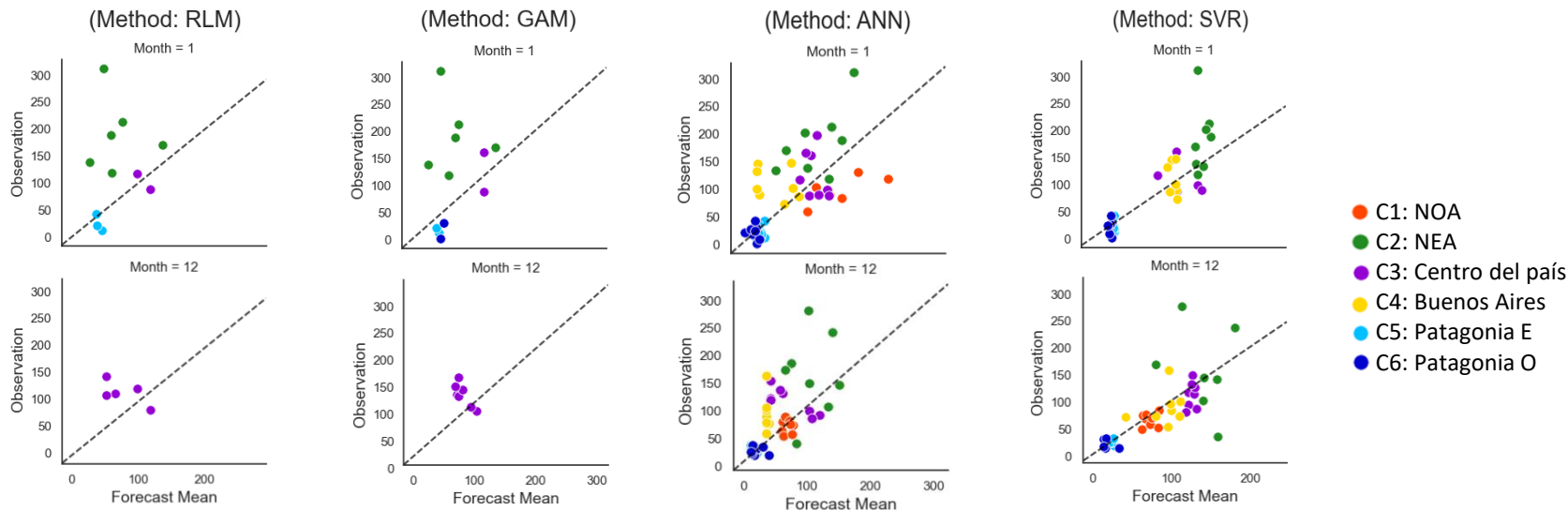


Fig. 9: Diagrama de dispersión del valor observado versus valor pronosticado junto con la línea de identidad (línea punteada) para las diferentes metodologías, para enero y diciembre a modo de ejemplo.

2 Verificación

3. RMSE del Ensamble y por metodología

- Centro del país (C3): abril a septiembre en C3
- Buenos Aires (C4): abril a julio
- Patagonia O (C6): mayo a octubre

- SVR presenta menor RMSE en la mayoría de las regiones, especialmente en Patagonia (C5 y C6)
- En invierno especialmente predominan ANN y SVR. En verano únicamente en diciembre, no hay predominancia clara en enero y febrero (no mostrado)

C1: NOA
C2: NEA
C3: Centro del país
C4: Buenos Aires
C5: Patagonia E
C6: Patagonia O

RMSE Pronóstico por Ensamble

Periodo de verificación: 2016-2023

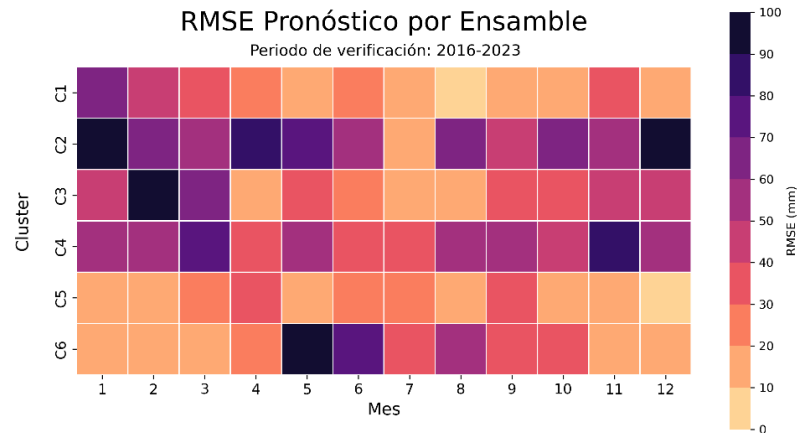


Fig. 10: Mapa de calor del RMSE del ensamble por clusters y mes.

Frequency of methods with lowest RMSE by cluster

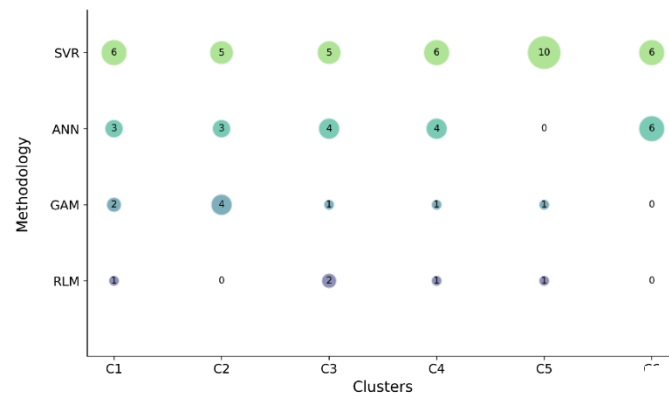


Fig. 11: Metodología con el menor RMSE por mes.

3 Pronóstico Operativo

Comunicación al usuario



Defensa Civil

Planificación de tareas, preparación de equipamiento, coordinación entre municipios y entre gobiernos locales con gobierno provincial, medidas de prevención en espacios públicos y arbolado, programación de tareas de limpieza de arroyos.



Actividades Agropecuarias

Preparación de suelo y la siembra, monitoreo continuo del desarrollo del cultivo (germinación, crecimiento y madurez). Procesamiento y almacenamiento hasta su comercialización en el mercado nacional o internacional.



Recursos Hídricos y Energía

Planificación de actividades y gestión de recursos, planificación en el sector hidro-energético tanto de la generación como de la demanda de energía.

Conclusiones

- La inclusión de técnicas de aprendizaje supervisado en los modelos de pronóstico ha demostrado mejoras significativas, que varían según la región y el mes analizado.
- El ensamble de metodologías lineales, no lineales y de machine learning ha mostrado pronósticos cercanos a los valores observados, con resultados particularmente destacados en los meses de marzo, noviembre y de julio a septiembre.

Conclusiones

- En el Litoral, donde las precipitaciones son más abundantes, se observa una mayor dispersión en los pronósticos entre octubre y diciembre, aunque las técnicas de aprendizaje supervisado y la inclusión de procesos no lineales parecen mejorar los resultados en esta región.
- En el centro del país y la región de Buenos Aires, el ensamble muestra mejores resultados durante el semestre frío, cuando las precipitaciones son menores.

Conclusiones

- El análisis del error por metodologías reveló que la técnica SVR ofrece los mejores resultados en la mayoría de los meses y regiones, con un impacto particularmente notable en la Patagonia.
- En el futuro se explorarán otras métricas que evalúen la precisión y confiabilidad de los pronósticos, con el objetivo de que puedan ser utilizados de manera operativa en el futuro cercano.



Consultas

egarbarini@smn.gov.ar



Servicio
Meteorológico
Nacional
Argentina



Ministerio
de Defensa
República Argentina

Dorrego 4019 (C1425GBE) Buenos Aires . Argentina

Tel: (+54 11) 5167-6767 . smn@smn.gov.ar

www.smn.gov.ar

