

Pronóstico de la incidencia semanal de depresión en México: marco multimodelo con selección auditable por serie

Javier Augusto Rebull-Saucedo¹ · Juan Carlos Pérez-Nava¹ · Luis Gerardo Sánchez-Salazar¹
Grettel Barceló-Alonso¹ · Ruth Pérez-Hernández²

¹ Tecnológico de Monterrey, Monterrey, México

² Instituto Mexicano del Seguro Social, OOAD Guerrero, México
Investigación independiente; las afiliaciones no implican aval institucional.

*Cuando las cifras históricas se enfrentan
a los pronósticos registrados de 2026*

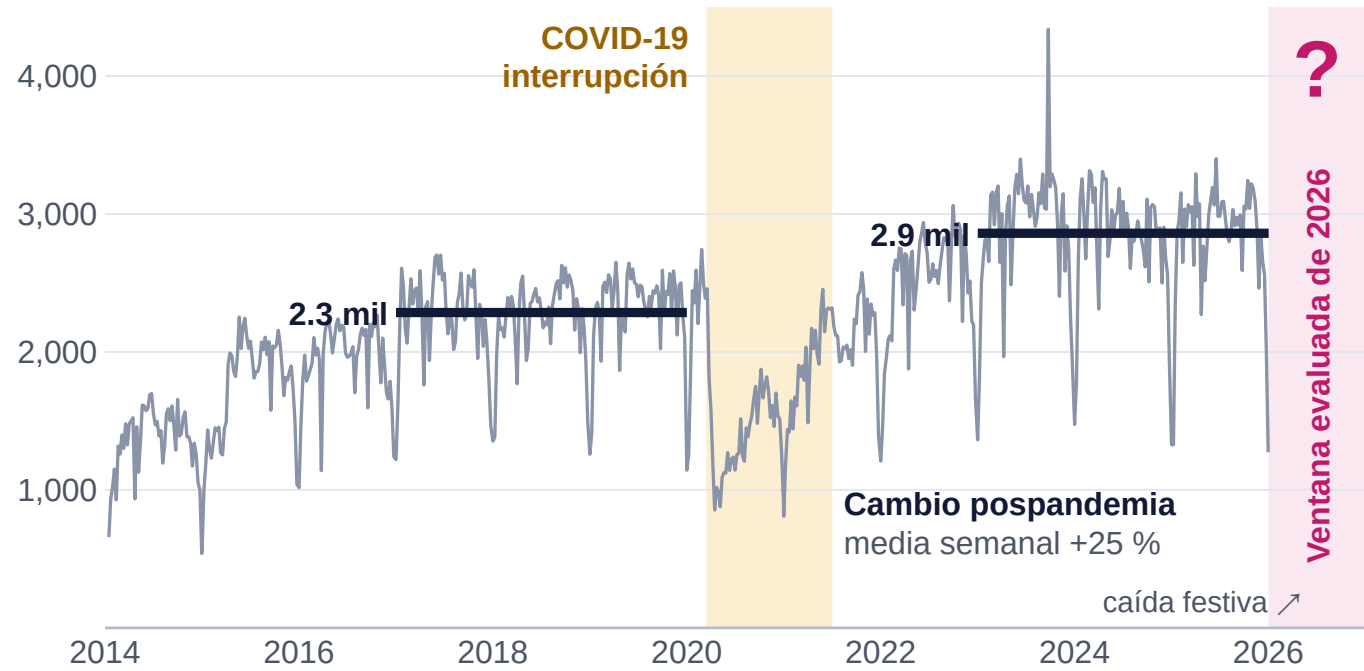


Tecnológico
de Monterrey



La notificación semanal describe el pasado

Casos semanales notificados, nacional CIE-10 F32, SINAVE 2014–2025



~150 mil **74 %**
casos al año **son mujeres**
2023–2025 total anual por sexo



Fuente: boletines del SINAVE [1]; conteos notificados. Líneas: medias semanales 2017–19 y 2023–25. La curva omite 2014-S02 y 2016-S20 (marcadas como parciales).

Cuatro candidatos por serie, una regla auditable



SINAVE boletines

Conteos semanales
CIE-10 F32, 2014–2025
(626 semanas)

111 objetivos: 32
estados, 4 regiones,
nacional; general,
mujeres, hombres

99 notificados
directamente



Cuatro candidatos

- **DeepAR** [3]
red recurrente probabilística
- **Prophet** [2]
tendencia + estacionalidad
- **Ensemble** [4]
Prophet + XGBoost
- **Stacking** [5, 6]
Prophet, ETS, LightGBM



Regla de selección determinista

- 1 **Menor sMAPE histórico**
- 2 **Dentro del 5 % relativo: menor MASE**
- 3 **Empate en MASE: menor RMSE**



Pronóstico registrado

52 semanas
adelante



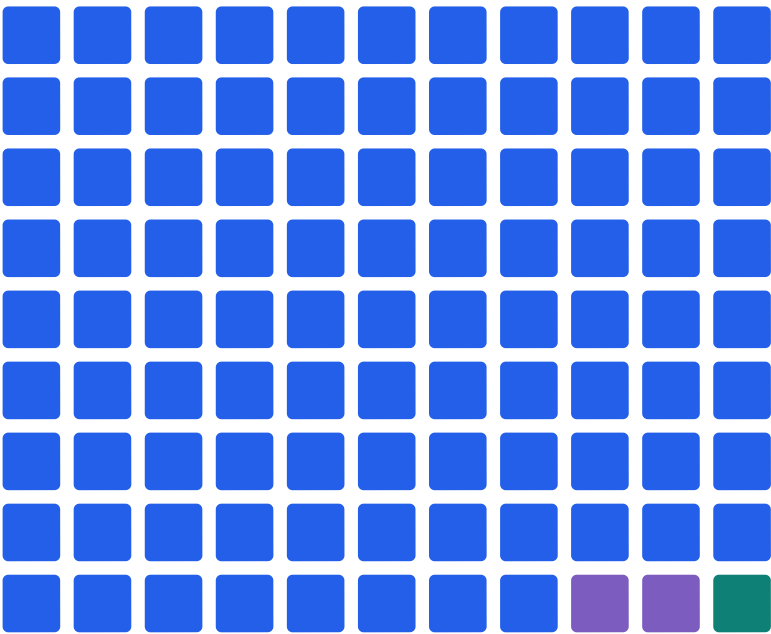
Bitácora

qué modelo y con
qué evidencia

Evidencia de selección: cifras históricas por modelo (tabla 1 del artículo), con protocolos distintos. MASE como en [7]. Respaldo para series escasas: 0 de 111.

Asignaciones iniciales registradas

99 objetivos evaluados un cuadro cada uno · asignación registrada



-  **DeepAR 96**
los 32 estados × 3 estratos
-  **Ensemble 2**
nacional, general y mujeres
-  **Prophet 1**
nacional, hombres

Objetivos estatales con DeepAR

96 / 96

elegidos con cifras históricas de selección

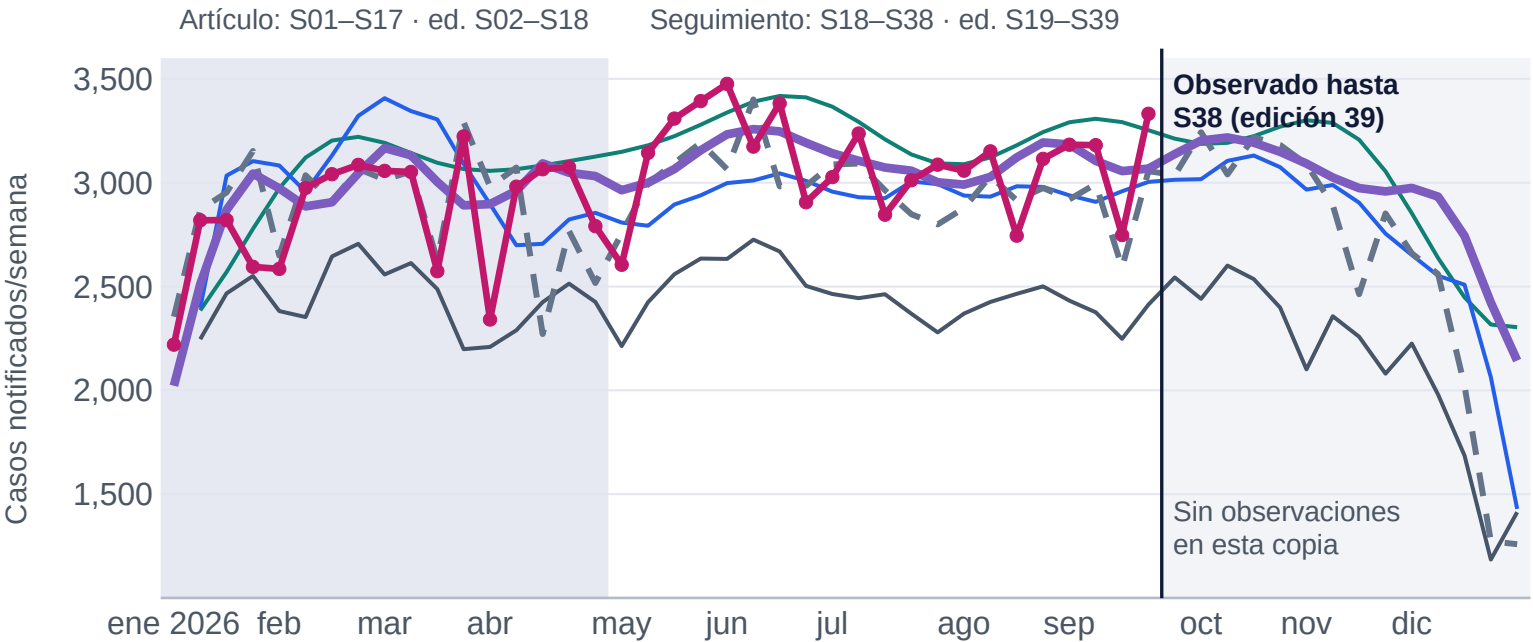
Nota: Los protocolos históricos difieren. El código estatal de DeepAR evalúa ajuste en entrenamiento.

Regla de selección aplicada a las cifras históricas de la tabla 1 del artículo (107 de 111 objetivos, regiones incluidas, a DeepAR). Cifras históricas en el respaldo B8.

Trayectorias registradas de 2026 y seguimiento

Nacional, estrato general Prophet, DeepAR y Stacking empiezan en S02

— Observado — Ensemble — Ingenuo — Prophet — DeepAR — Stacking



99 objetivos, ediciones S19–S39

SMAPE mediano · bajo el ingenuo

Prophet	24.2	72/99
Ensemble	25.7	70/99
DeepAR	31.7	57/99
Stacking	31.4	48/99
Archivada	31.7	60/99
Ingenuo	28.6	

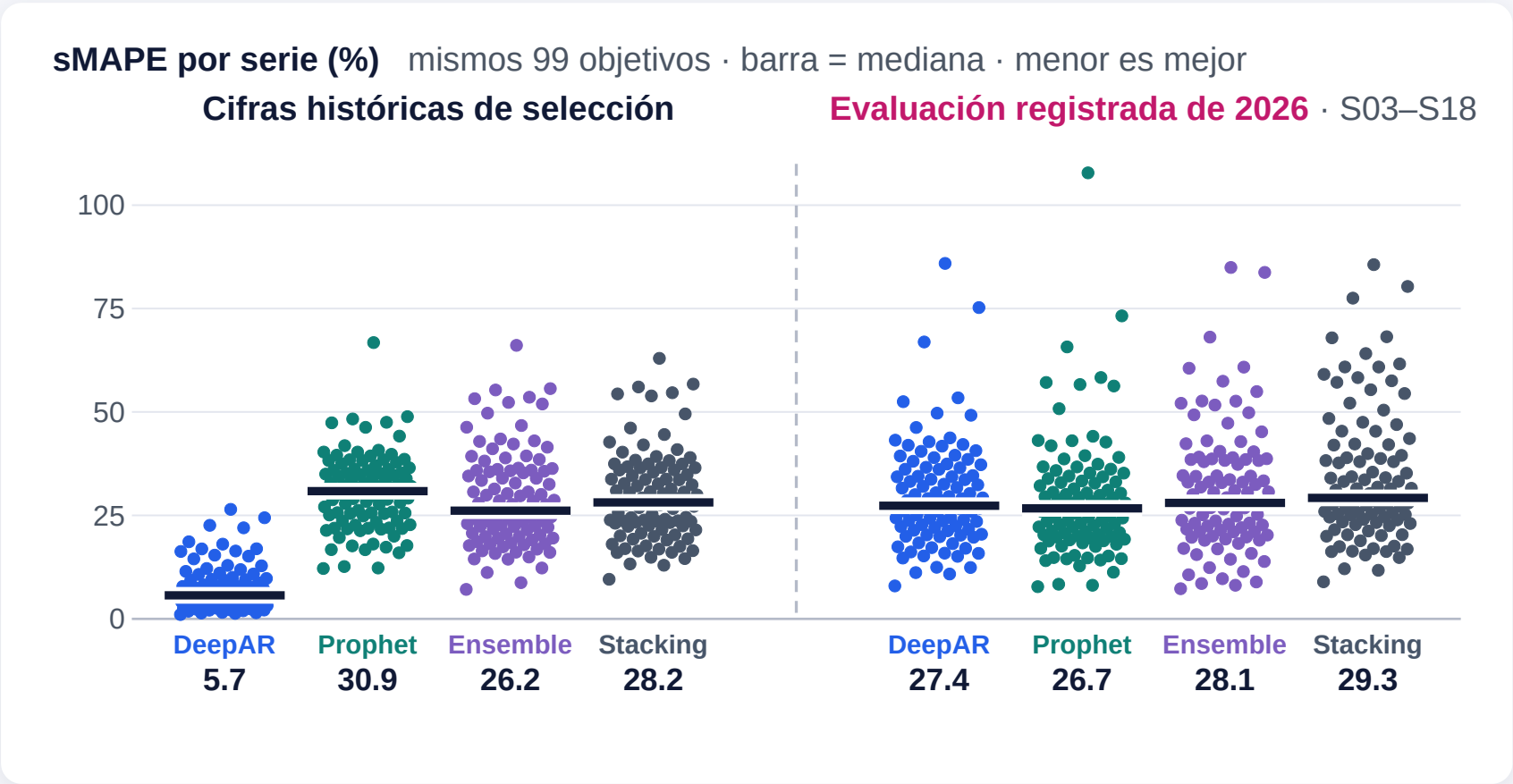
Nacional, ediciones S19–S39

Menor: Ensemble

Ensemble	5.64 %
Ingenuo 52 sem.	6.36 %

Pronósticos registrados · evaluación retrospectiva. Curvas: un objetivo nacional; tabla: 99 objetivos. Corte: edición 39. Descriptivo. Archivada = asignación registrada.

Los órdenes histórico y de 2026 difieren



Los protocolos históricos difieren.
Comparación descriptiva en los mismos 99 objetivos.

Rango de medianas 2026

< 3 puntos

Menor mediana: Prophet

Ningún modelo gana en todas las series

35	27	19	18
----	----	----	----

objetivos ganados en 2026: Prophet, DeepAR, Ensemble, Stacking

Panel derecho y tarjetas: errores registrados, ediciones S03–S18 (ventana del artículo). Post hoc y exploratorio: objetivos superpuestos; pruebas en B2. Figura completa en B9.

Sin DeepAR, los dos órdenes difieren

sMAPE mediano (%) Ensemble y Prophet · 99 series · menor es mejor

Histórica ventanas reservadas



Evaluación registrada de 2026 S03–S18



Mismo modelo de menor error, histórico y 2026

Prophet, Ensemble, Stacking · 99 series

34 % observado
34 de 99 series

31 % referencia
según las frecuencias de victoria

Referencia descriptiva, no una prueba formal de azar.

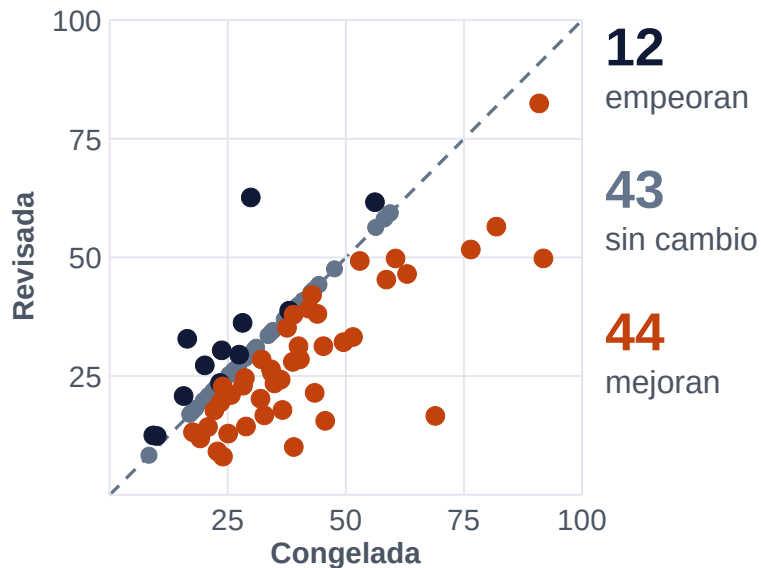
Análisis adicional al artículo. Órdenes puntuales; post hoc y exploratorio. Promediar los tres objetivos de cada estado debilita la inversión (respaldo B2).

La revisión bajó la mediana en 7 semanas reconstruidas

S03–S11 reseleccionar: misma regla, solo semanas previas

S12–S18 evaluación: excluidas de la reselección

Cada punto es una serie SMAPE, S12–S18



Políticas SMAPE mediano, S12–S18

Asignación congelada **32.7**

Reselección global **32.7**

Revisión por serie 28.5

Comparadores

Solo Prophet **27.6**

Conjunto sin DeepAR **26.4**

-4.2 puntos diferencia de medianas de política

Lectura

En siete semanas reconstruidas, la revisión bajó la mediana de 32.7 % a 28.5 %.

Las alternativas simples tuvieron medianas menores.

Revisada frente a congelada: 44 mejoran, 12 empeoran, 43 iguales. La mediana conjunta fue menor en los nueve cortes (S06–S14); en S11 fue menor en cada estrato. Exploratorio.

CONCLUSIONES

La selección es una señal, no una garantía

Un marco de selección auditable por serie, evaluado con pronósticos registrados de 2026.

1

Los órdenes histórico y de 2026 difieren

La selección histórica favoreció a DeepAR. Prophet tuvo el menor error mediano en el panel registrado.

2

La revisión bajó la mediana

Reconstrucción de siete semanas: revisión 32.7 → 28.5; conjunto sin DeepAR 26.4.

3

Cómo va el año

Seguimiento de 21 semanas: Ensemble nacional 5.64, ingenuo 6.36; mediana archivada 31.7, ingenuo 28.6.

GRACIAS

¿Preguntas?

La selección es una señal, no una garantía.

Forecasting Weekly Depression Incidence in Mexico: A Multi-Model Framework with Auditable Per-Series Model Selection

En prensa en Advances in Soft Computing (MICAI 2026), Springer LNCS 17149



rebull@exatec.tec.mx



[github.com/EpiForecastMX/
per-series-model-selection2026](https://github.com/EpiForecastMX/per-series-model-selection2026)



epiforecast.mx/micai2026
seguimiento, datos y diapositivas

Referencias

- [1] Dirección General de Epidemiología, Secretaría de Salud: SINAVE Boletín Epidemiológico (2026).
- [2] Taylor, S.J., Letham, B.: Forecasting at scale. Am. Stat. 72(1), 37–45 (2018).
- [3] Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T.: DeepAR: probabilistic forecasting with autoregressive recurrent networks. Int. J. Forecast. 36(3), 1181–1191 (2020).
- [4] Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: KDD, pp. 785–794 (2016).
- [5] Ke, G., et al.: LightGBM: a highly efficient gradient boosting decision tree. In: NeurIPS 30, pp. 3146–3154 (2017).
- [6] Wolpert, D.H.: Stacked generalization. Neural Netw. 5(2), 241–259 (1992).
- [7] Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. Int. J. Forecast. 22(4), 679–688 (2006).
- [8] Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: long-term forecasting with transformers. In: ICLR (2023).
- [9] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time Series Analysis: Forecasting and Control, 5th edn. Wiley (2015).

Este trabajo. Rebull-Saucedo, J.A., Pérez-Nava, J.C., Sánchez-Salazar, L.G., Barceló-Alonso, G., Pérez-Hernández, R.: Forecasting weekly depression incidence in Mexico: a multi-model framework with auditable per-series model selection. En: Advances in Soft Computing, MICAI 2026, LNCS 17149, Springer (en prensa).

La regla de selección, con precisión

Por serie, con el SMAPE, MASE y RMSE históricos registrados de cada modelo:

- 1 Si el segundo menor SMAPE es mayor que 1.05 veces el menor, se elige el modelo de menor SMAPE.
 - 2 Si no, entre los modelos con $\text{SMAPE} \leq 1.05 \times \text{el mejor}$, elegir el menor MASE.
 - 3 Si el MASE empata, se elige el menor RMSE.
- + Menos de 5 casos en las 52 semanas anteriores al origen del pronóstico: modelo regional (solo datos pasados; nunca se activó, 0 de 111).

Cada elección y sus métricas se guardan con el pronóstico: la bitácora auditable.

Evidencia de selección

Evaluaciones históricas por modelo (tabla 1 del artículo), sin un protocolo común.

Prophet, Ensemble, Stacking: ventanas reservadas, con protocolos distintos.

El código histórico de DeepAR estatal omite la VC, entrena con la serie completa y califica ese ajuste (commit 60d6ea0f; cambiado en 9fdd6dca). Las ejecuciones publicadas no están vinculadas de forma independiente.

El artículo reporta datos hasta 2025-S53; las fechas de ejecución no están verificadas de forma independiente.

Pruebas estadísticas e incertidumbre

Comparación	99 objetivos: nominal / Holm	32 estados: nominal
Histórico, Ensemble frente a Prophet	0.020 / 0.060	0.15
2026, Prophet frente a Ensemble	0.007 / 0.035	0.06
2026, Prophet frente a DeepAR	0.070 / 0.14	0.24
Revisada frente a congelada, corte S11	< 0.001 / < 0.001	< 0.001

Exploratorio y post hoc; los objetivos se superponen y Holm no elimina la dependencia. No significativo ≠ equivalente. Wilcoxon bilateral, zsplit; Holm dentro de seis pares de modelos (nueve para la revisión).

Nacional, S03–S18, recalculado

Devianza de Poisson media: Ensemble 25.7, Prophet 32.3, DeepAR 46.0, Stacking 97.8.

Solo reportado en el artículo

Pruebas de Diebold–Mariano; la banda del 80 % cubrió 10 de 17 semanas. Entradas no publicadas.

Limitaciones

Un solo padecimiento

Solo depresión (F32); no se probó la transferencia.

Búsqueda desigual

Prophet tuvo la búsqueda más amplia; efecto no probado.

Procedencia

Fechas de ejecución de los pronósticos no verificadas de forma independiente.

Incidencia notificada

Los modelos aprenden juntas la notificación y la enfermedad.

Evidencia de ajuste

DeepAR estatal histórico: ajuste en entrenamiento; procedencia de ejecuciones no verificada.

Evaluación de parte del año

Ventanas del artículo y 21 períodos de seguimiento (hasta S38); falta el año completo.

Incertidumbre

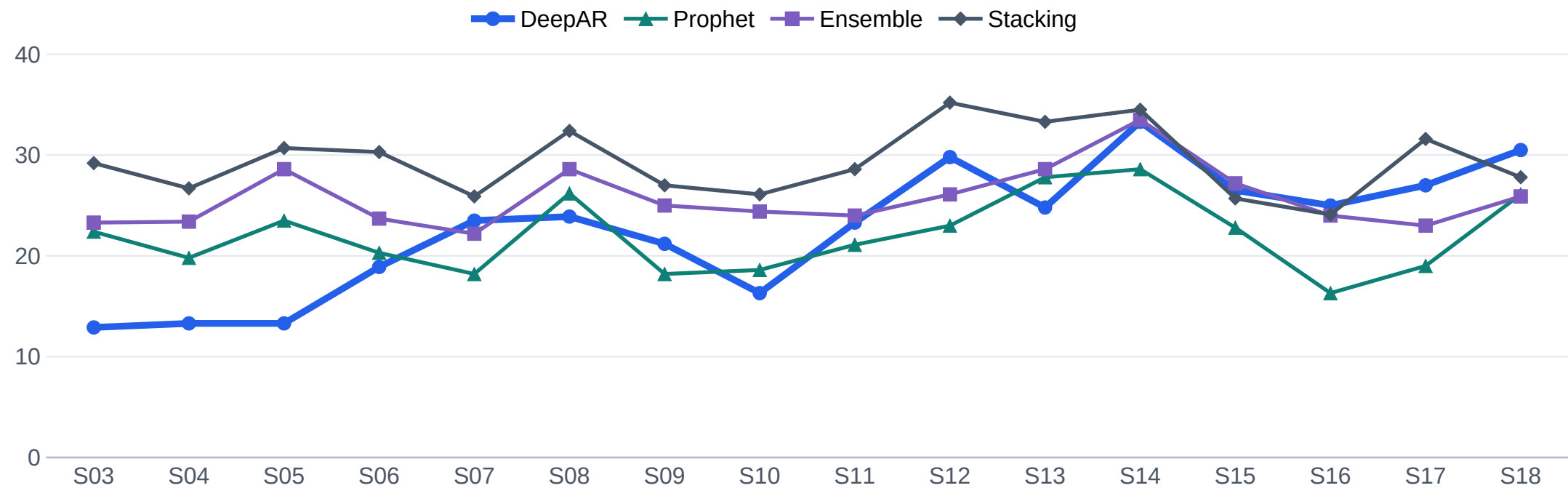
Cobertura reportada en el artículo: 58.8 % para la banda del 80 %; revisar no la calibra.

Sin validación externa

Solo SINAVE y México; un prototipo de investigación.

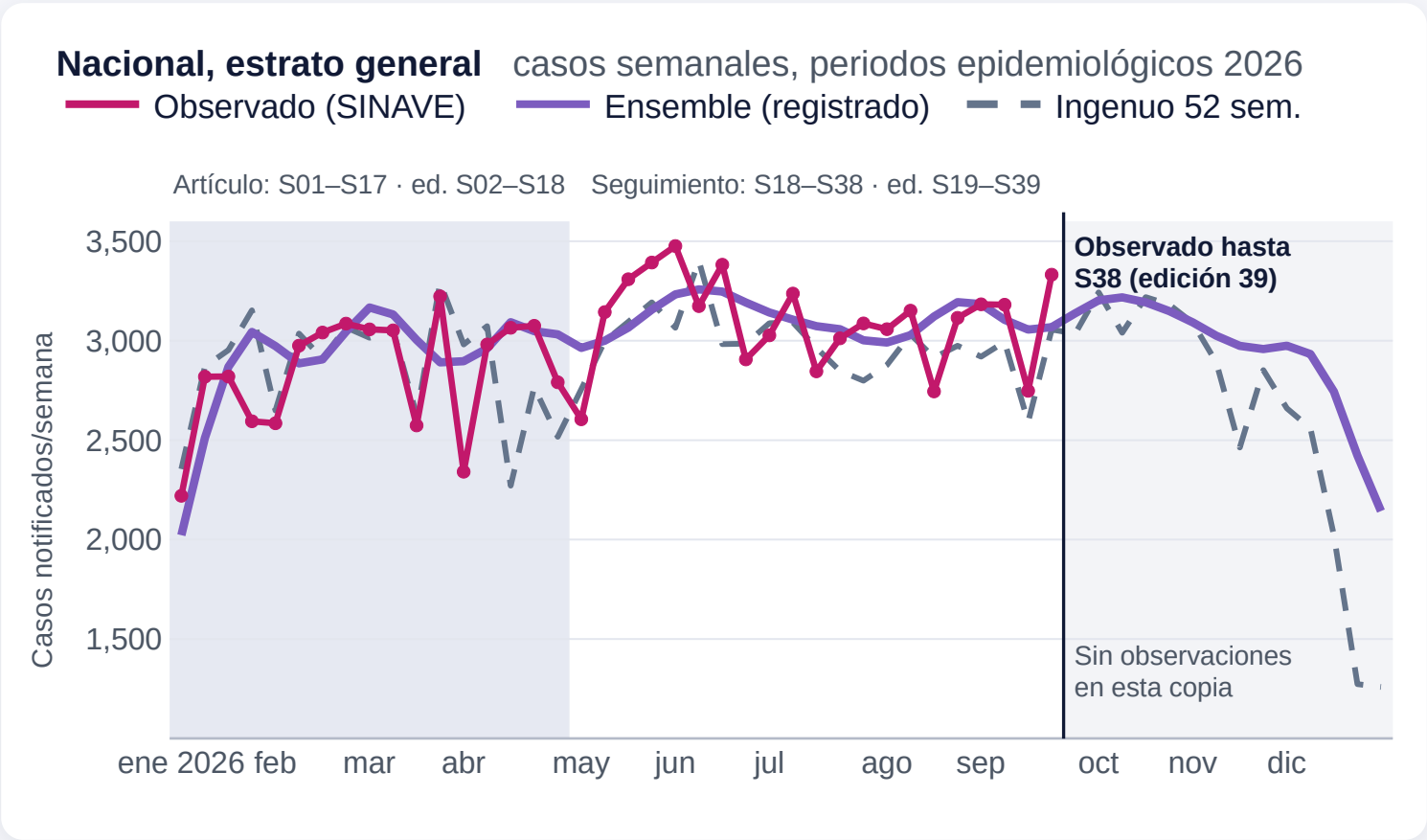
DeepAR lidera las primeras cuatro ediciones (S03–S06)

sMAPE semanal mediano (%) en las 99 series ediciones del boletín 2026 · menor es mejor



La revisión elige entre trayectorias congeladas de 52 semanas; el horizonte puede contribuir a su ganancia (no aislado). Recalculado con los archivos públicos.

Pronóstico nacional frente al ingenuo de 52 semanas



Artículo sMAPE (%)
ediciones S02–S18 · **Menor: ingenuo**

Ensemble **7.40 %**

Ingenuo 52 sem. **7.33 %**

Seguimiento sMAPE (%)
ediciones S19–S39 · **Menor: Ensemble**

Ensemble **5.64 %**

Ingenuo 52 sem. **6.36 %**

**Periodos del pronóstico
asignados por regla; origen de
ejecución no verificado.**

sMAPE: menor es mejor. Desviación acumulada, Ensemble / ingenuo: artículo +2.45 % / +0.6 %; seguimiento +0.21 % / –3.81 %. Corte: edición 39. Descriptivo.

Cada modelo frente al ingenuo, por ventana y sexo

sMAPE mediano (%) y objetivos que superan al ingenuo, de 33 por estrato · menor es mejor

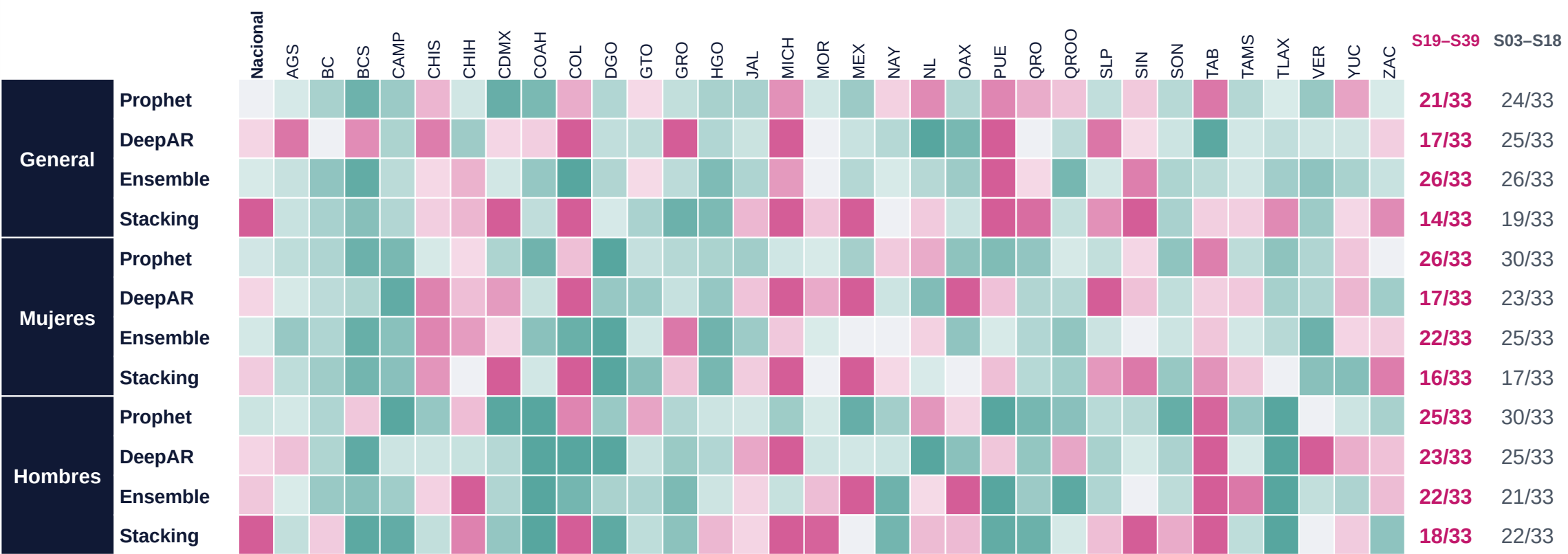
	Ventana del artículo · ediciones S03–S18			Seguimiento · ediciones S19–S39		
Estimador	General	Mujeres	Hombres	General	Mujeres	Hombres
Prophet	23.8 · 24	24.4 · 30	30.4 · 30	21.3 · 21	22.7 · 26	30.0 · 25
DeepAR	25.9 · 25	26.0 · 23	32.4 · 25	23.3 · 17	27.1 · 17	35.2 · 23
Ensemble	26.7 · 26	23.8 · 25	38.0 · 21	21.8 · 26	25.7 · 22	36.3 · 22
Stacking	29.0 · 19	27.9 · 17	34.3 · 22	27.8+ · 14	30.6+ · 16	39.7+ · 18
Asignación archivada	25.9 · 26	26.0 · 24	32.4 · 25	23.3 · 18	27.1 · 18	35.2 · 24
Ingenuo 52 sem.	31.2	30.6	40.1	23.7	27.8	39.0

Celda: sMAPE mediano en 33 objetivos · objetivos con menor sMAPE que el ingenuo. + = mediana por encima del ingenuo; negritas = menor mediana. Asignación archivada = DeepAR en todos los estados. Descriptivo; objetivos superpuestos.

Cómo va el año: cada modelo frente al ingenuo

Seguimiento, ediciones S19–S39 · celda = sMAPE del modelo menos sMAPE del ingenuo, un objetivo cada una

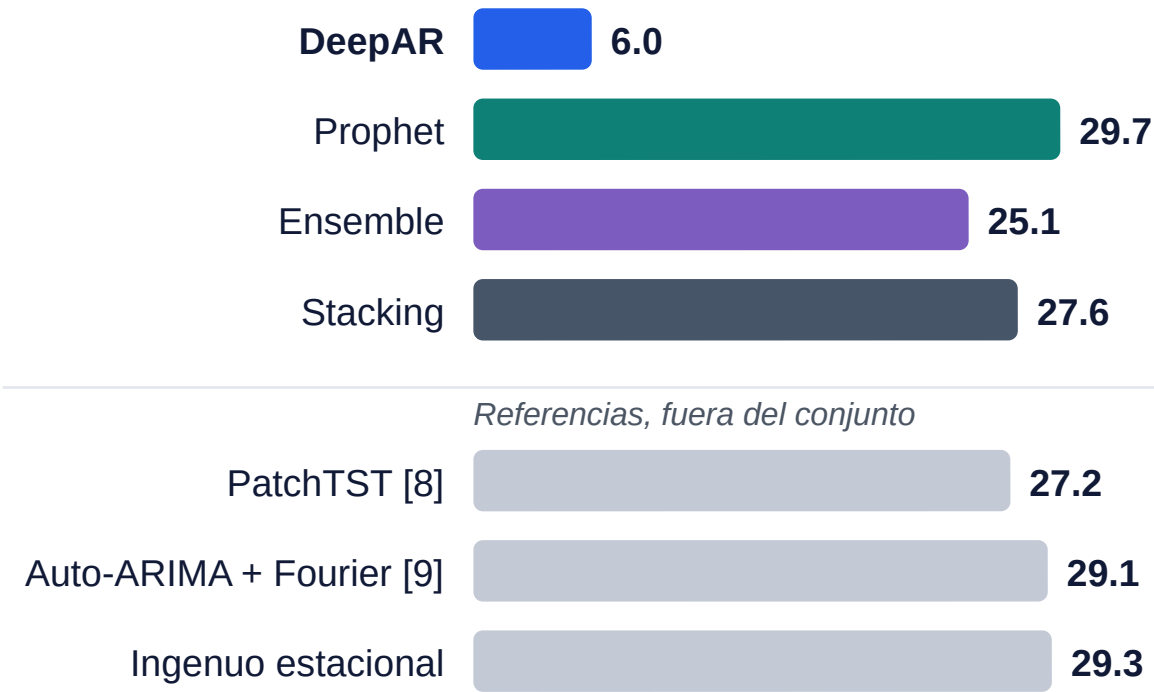
Objetivos que superan al ingenuo



Verde azulado: menor error que el ingenuo (mejor). Magenta: mayor. Gris: menos de 0.5 pts, no necesariamente empate. Asignación archivada = DeepAR en todos los estados. Descriptivo; objetivos superpuestos. Matriz completa para descarga; la lectura por estado requiere zoom.

Cifras históricas de selección (tabla 1 del artículo)

sMAPE histórico mediano (%) Tabla 1 · 111 objetivos · menor es mejor



Series que la regla asigna a DeepAR

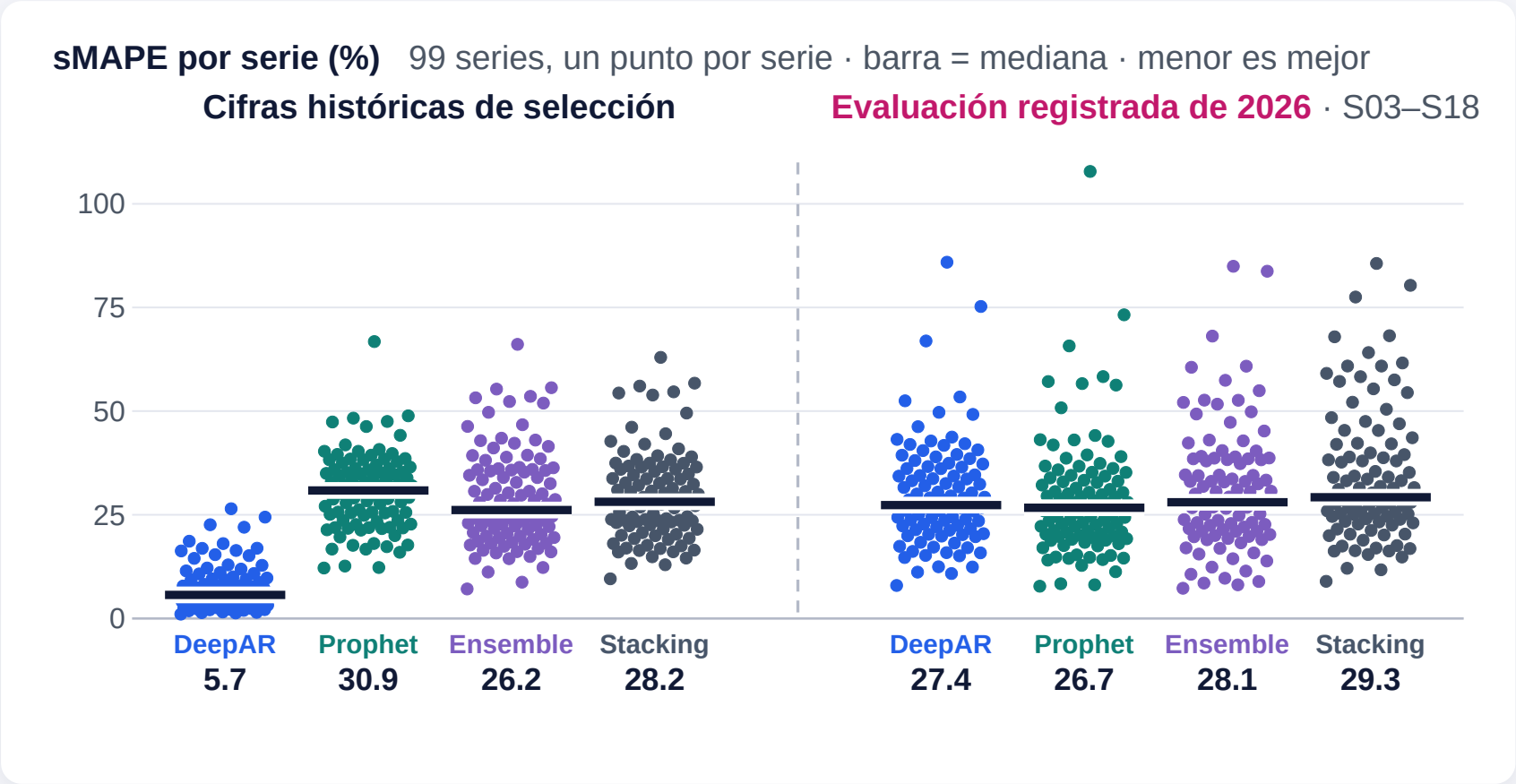
107 / 111

96.4 %, incluidas todas las series estatales

Cómo leer el 6.0: en el código histórico, las cifras estatales de DeepAR reflejan ajuste en entrenamiento, no error de pronóstico.

El 6.0 % de DeepAR es la mediana de los 111 objetivos. Las cifras históricas de selección no garantizan la exactitud de 2026.

Los órdenes históricos y de 2026 difieren



Diferencia histórica
≈20 puntos
Rango de medianas 2026
< 3 puntos

Ningún modelo gana en todas las series

35	27	19	18
----	----	----	----

series ganadas en 2026: Prophet, DeepAR, Ensemble, Stacking

Los protocolos históricos difieren entre modelos. Análisis post hoc y exploratorio: los objetivos se superponen; pruebas pareadas en el respaldo B2.