

Makarova, E.; Portabella, M.; Stoffelen, A. Corrección de sesgos en los vientos de ERA5 mediante aprendizaje automático

Corrección de sesgos en los vientos de ERA5 mediante aprendizaje automático

Makarova, Evgeniia ¹ Portabella, Marcos ¹ Stoffelen, Ad ²

¹ Barcelona Expert Center, ICM - CSIC, España

² KNMI

ORCID: Makarova [0000-0003-1418-8385](https://orcid.org/0000-0003-1418-8385) Portabella [0000-0002-9972-9090](https://orcid.org/0000-0002-9972-9090) Stoffelen [0000-0002-4018-4073](https://orcid.org/0000-0002-4018-4073)

Correspondencia: makarova@icm.csic.es portabella@icm.csic.es ad.stoffelen@knmi.nl

RESUMEN

Los vientos oceánicos superficiales precisos son fundamentales para la modelización oceánica; sin embargo, reanálisis ampliamente utilizadas, como el Reanálisis v5 del ECMWF (ERA5), presentan sesgos locales y persistentes. Este estudio evalúa métodos de aprendizaje automático (ML) para corregir dichos sesgos mediante redes neuronales totalmente conectadas (FCNNs), entrenadas con cinco años de variables atmosféricas y oceánicas de ERA5, junto con datos de corrientes superficiales del Servicio de Vigilancia Marina Copernicus (CMEMS). Los resultados muestran que las correcciones basadas en ML reducen significativamente los sesgos de viento en ERA5: a escala global, la reducción de varianza del error alcanza aproximadamente el 13% frente a ASCAT y el 9% frente a HSCAT, con las mayores mejoras (hasta un 16%) observadas en las regiones extratropicales. Los modelos mensuales superan a los modelos anuales, especialmente en los trópicos, donde la variabilidad estacional influye de manera notable en los patrones de sesgo. Aunque persisten algunas discrepancias al validarse con dispersómetros en banda Ku, los campos corregidos reducen sustancialmente las estructuras de sesgo a gran escala.

Palabras clave: *redes neuronales, dispersómetros, vientos equivalentes a estrés, sesgos de NWP*

Fecha de recepción: 3 febrero 2026 · Fecha de aceptación: 17 marzo 2026


Corrección de sesgos en los vientos de ERA5 mediante aprendizaje automático

Makarova, Evgeniia ⁽¹⁾, Portabella, Marcos ⁽¹⁾, Stoffelen, Ad ⁽²⁾

⁽¹⁾ Barcelona Expert Center, ICM - CSIC, España.

 0000-0003-1418-8385, makarova@icm.csic.es;  0000-0002-9972-9090, portabella@icm.csic.es.

⁽³⁾ KNMI

 0000-0002-4018-4073, ad.stoffelen@knmi.nl.

Resumen: Los vientos oceánicos superficiales precisos son fundamentales para la modelización oceánica; sin embargo, reanálisis ampliamente utilizadas, como el Reanálisis v5 (ERA5) del Centro Europeo de Previsiones Meteorológicas a Plazo Medio (ECMWF), presentan sesgos locales y persistentes. Este estudio evalúa métodos de aprendizaje automático (ML) para corregir dichos sesgos mediante redes neuronales totalmente conectadas (FCNNs), entrenadas con cinco años de variables atmosféricas y oceánicas de ERA5, junto con datos de corrientes superficiales del Servicio de Vigilancia Marina Copernicus (CMEMS).

Los resultados muestran que las correcciones basadas en ML reducen significativamente los sesgos de viento en ERA5: a escala global, la reducción de varianza del error alcanza aproximadamente el 13% frente a ASCAT y el 9% frente a HSCAT, con las mayores mejoras (hasta un 16%) observadas en las regiones extra-tropicales. Los modelos mensuales superan a los modelos anuales, especialmente en los trópicos, donde la variabilidad estacional influye de manera notable en los patrones de sesgo. Aunque persisten algunas discrepancias al validarse con dispersómetros en banda Ku, los campos corregidos reducen sustancialmente las estructuras de sesgo a gran escala.

Palabras clave: redes neuronales, dispersómetros, vientos equivalentes a estrés, sesgos de NWP

Bias correction of ERA5 winds with Machine Learning

Abstract: *Accurate ocean surface winds are essential for ocean modeling, yet widely used reanalyses such as ECMWF Reanalysis v5 (ERA5) exhibit local and persistent biases. This study assesses machine-learning (ML) methods to correct these biases using fully connected neural networks (FCNNs) trained on five years of ERA5 atmospheric and oceanic variables, combined with surface current data from Copernicus Marine Environment Monitoring Service (CMEMS).*

Results demonstrate that ML-based corrections significantly reduce ERA5 wind biases: on a global scale, error variance reductions reached approximately 13% against ASCAT and 9% against HSCAT, with the highest improvements (up to 16%) observed in extra-tropical regions. Monthly models outperformed all-year models, particularly in the tropics, where seasonal variability strongly influences bias patterns. Although residual discrepancies remain when validated against Ku-band scatterometers, the corrected outputs substantially mitigate large-scale bias structures.

Keywords: *neural networks, scatterometers, stress-equivalent winds, NWP biases*

1. INTRODUCCIÓN

La predicción precisa de los campos de viento oceánico en superficie es un elemento esencial para la modelización de la circulación oceánica, de oleaje y para las numerosas aplicaciones climáticas. Sin embargo, los sistemas de predicción numérica del tiempo (NWP), como el Integrated Forecasting System (IFS) del ECMWF, presentan sesgos recurrentes y espacialmente estructurados cuando se comparan con medidas de dispersómetros satelitales.

Estos sesgos afectan tanto a la dirección como a la intensidad del viento y están asociados a procesos de interacción océano-atmósfera que no están completamente representados en el modelo, como los efectos de los gradientes de temperatura superficial del mar (SST) y de las corrientes oceánicas, así como la representación de la convección húmeda (Sandu *et al.*, 2020).

Estudios previos han documentado rotaciones sistemáticas de los vientos superficiales y patrones

persistentes de los sesgos en el reanálisis ERA5, entre ellos un exceso de vientos del oeste en latitudes medias y una convergencia insuficiente hacia la Zona de Convergencia Intertropical (ZCIT) (Belmonte Rivas & Stoffelen, 2019). Además, estos errores reaparecen rápidamente después de la asimilación de vientos de dispersómetro, lo que indica que las causas dinámicas subyacentes permanecen sin corregir.

Diversos productos corregidos, como ERA* o los campos WIND TAC del CMEMS, reducen los sesgos persistentes aplicando correcciones derivadas de diferencias entre modelos y dispersómetros (Trindade *et al.*, 2020; CMEMS, 2023). Sin embargo, su rendimiento depende fuertemente de la disponibilidad espacial y temporal de observaciones de dispersómetro, lo que limita su aplicabilidad en predicciones operacionales, especialmente en regiones o periodos con poca cobertura satelital.

Para superar estas limitaciones, trabajos recientes han analizado enfoques de aprendizaje automático capaces de aprender relaciones sistemáticas entre variables atmosféricas y oceánicas y los errores de viento en ERA5 (Makarova *et al.*, 2025). Siguiendo esta línea, el presente estudio investiga el uso de redes neuronales totalmente conectadas (FCNNs) entrenadas con cinco años de datos para generar correcciones de los sesgos que pueden aplicarse incluso cuando no se dispone de observaciones de dispersómetros.

2. MATERIAL Y METODOS

2.1. Conjuntos de datos

Los modelos de aprendizaje automático se entrenaron utilizando cinco años de datos atmosféricos y oceánicos (2019–2022, 2024), reservando el año 2018 para la validación y 2023 para las pruebas independientes. Los objetivos de entrenamiento consisten en las diferencias entre los vientos del dispersómetro Advanced scatterometer (ASCAT) en banda C y los vientos equivalentes a estrés a 10 metros de ERA5 (U10S). Los vientos ASCAT de la serie MetOp proporcionan vectores de viento superficiales de alta calidad en todas las condiciones meteorológicas. En este trabajo se usan los productos de nivel 2 (L2) en malla regular con una resolución espacial de 12,5 km.

Las variables de entrada se extrajeron de la reanálisis ERA5 (Hersbach *et al.*, 2020), incluyendo los componentes de U10S, la presión a nivel del mar, la temperatura de la superficie, la humedad específica, la temperatura superficial del mar (SST) y campos derivados como vorticidad, divergencia y los gradientes de SST. Para tener en cuenta los efectos de las corrientes en la superficie oceánica, se añadieron corrientes superficiales medias diarias procedentes del conjunto de datos CMEMS Global Total, Ekman and Geostrophic currents at the Surface and 15 m (GLOBCURRENT) (CMEMS, 2023). Todos los conjuntos de datos se colocaron espacial y temporalmente y se normalizaron para garantizar la coherencia entre las entradas atmosféricas, oceánicas y del dispersómetro.

Para la validación independiente, los vientos corregidos también se compararon con vientos de los dispersómetros HSCAT en banda Ku (productos L2 de 25 km) procedentes de los satélites Haiyang HY-2B y HY-2C, que proporcionan mediciones de viento en superficie con horarios locales de paso diferentes a los de las misiones ASCAT.

2.2. Metodología

Primero los conjuntos de datos de ERA5 y GLOBCURRENT se colocaron en el espacio y en el tiempo para ajustarlos a la malla L2 de ASCAT. Las derivadas, como la vorticidad, la divergencia y los gradientes de la SST se calcularon para coincidir con la misma malla de L2. Todas las variables de entrada se normalizaron y se hizo un submuestreo de los datos al 15 % con el fin de reducir el tiempo de entrenamiento.

La arquitectura de aprendizaje automático seleccionada para este estudio es una red neuronal *feed-forward* totalmente conectada, que en los trabajos previos ha mostrado resultados robustos (Makarova *et al.*, 2025). Tras evaluar varias configuraciones, la arquitectura elegida consiste en una red de cinco capas ocultas con el número de neuronas decrecientes progresivamente (1024–512–256–128–64), que ofrece un equilibrio óptimo entre precisión y capacidad de generalización sin un crecimiento excesivo del número de parámetros. Todas las redes se entrenaron utilizando el optimizador Adam y una función de pérdida basada en el error cuadrático medio (RMSE).

Se evaluaron dos estrategias de modelado. El modelo anual se entrenó con el conjunto completo de datos de cinco años, incorporando una codificación cíclica del día del año para representar la variabilidad estacional. Adicionalmente, los modelos mensuales se entrenaron con datos de una ventana temporal de cada mes específico (± 15 días alrededor de cada mes) para capturar mejor las estructuras de sesgo dependientes de la estación, especialmente en los trópicos, donde los errores del viento presentan una fuerte variabilidad estacional.

El rendimiento de los modelos se evaluó colocándolos los vientos corregidos con observaciones de ASCAT y HSCAT y calculando la diferencia vectorial cuadrática media (VRMS) (Ecuación 1), así como la reducción de la varianza del error asociada en relación con ERA5 (Ecuación 2).

$$VRMS = \frac{1}{N} \sum (u_{scat} - u_{model})^2 + (v_{scat} - v_{model})^2 \quad (1)$$

$$Err. var. red. = \frac{VRMS_{ERA5}^2 - VRMS_{NN}^2}{VRMS_{ERA5}^2} \quad (2)$$

Tabla 1. VRMS (m/s) medio anual de ERA5 frente a los dispersómetros ASCAT-B y -C y HSCAT-B y -C.

ERA5 VRMS	Global	Trópicos	Extra-Trópicos	Latitudes Altas
ASCAT-B	2.036	1.999	2.026	2.143
ASCAT-C	2.030	1.996	2.020	2.130
HSCAT-B	1.602	1.550	1.593	1.756
HSCAT-C	1.682	1.615	1.687	1.804

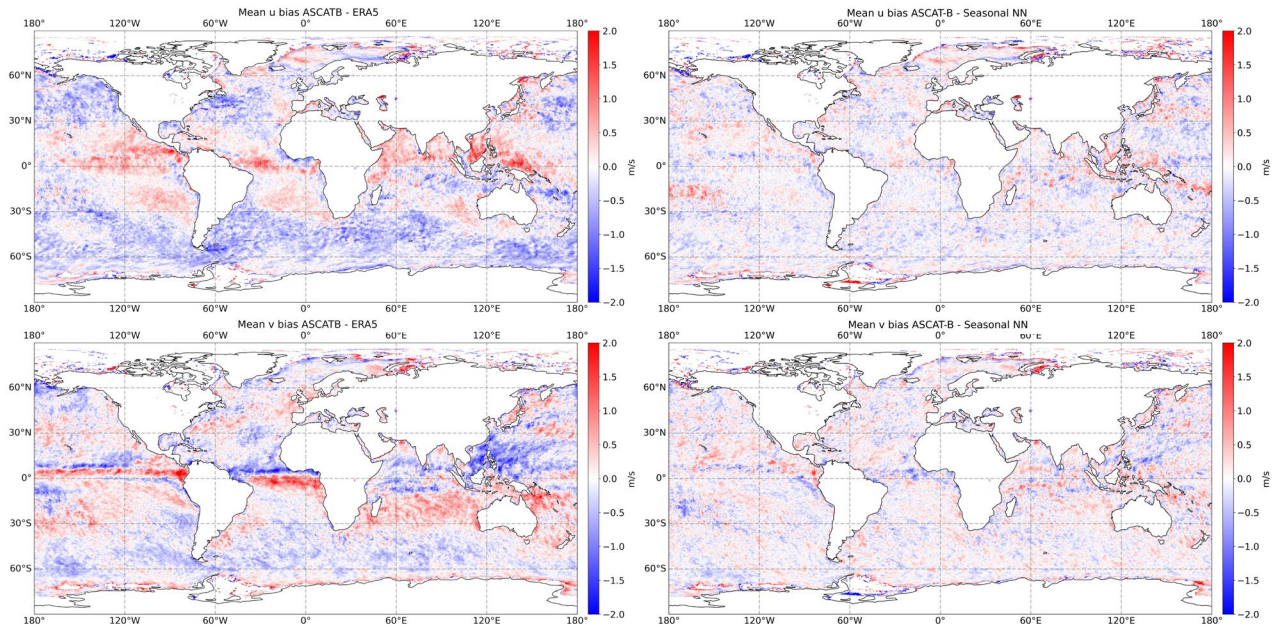


Figura 1. Sesgo medio (m/s) de ERA5 (izquierda) y de la salida corregida con red neuronal (derecha) frente a ASCAT-B para enero de 2023, tanto para la componente zonal (arriba) como para la componente meridional (abajo) de U10S.

3. RESULTADOS Y DISCUSSION

La evaluación de ERA5 frente a distintos dispersómetros revela que la magnitud y la estructura de los sesgos locales del viento dependen del instrumento utilizado para la validación. La Tabla 1 muestra que ERA5 presenta diferencias en VRMS globales más elevadas cuando se compara con los dispersómetros en banda C ASCAT-B y ASCAT-C ($\approx 2.0 \text{ m s}^{-1}$), mientras que los dispersómetros en banda Ku HSCAT-B y HSCAT-C presentan valores de VRMS más bajos ($\approx 1.6\text{--}1.7 \text{ m s}^{-1}$). Esta diferencia refleja la distinta resolución espacial, los procedimientos de control de calidad y las características de muestreo de ambos tipos de instrumentos, así como las diferentes condiciones meteorológicas retenidas tras aplicar el control de calidad, dado que los dispersómetros en banda Ku se ven sustancialmente más afectados por la contaminación debida a la precipitación que los dispersómetros en banda C.

En la Figura 1 (panel izquierdo), el sesgo medio de ERA5 frente a ASCAT-B muestra estructuras bien definidas a gran escala, incluyendo asimetrías zonales pronunciadas y un patrón de sesgo meridional asociado a la ZCIT. Los paneles derechos muestran los sesgos resultantes después de aplicar las correcciones del modelo

mensual, donde las estructuras de sesgo recurrentes desaparecen en gran medida, dejando un sesgo residual de carácter más granular.

La Tabla 2 resume el rendimiento de los modelos anuales y mensuales. Los modelos mensuales superan de forma sistemática al modelo anual, alcanzando reducciones globales de la varianza del error de aproximadamente el 13 % frente a ASCAT y el 9 % frente a HSCAT, en comparación con $\sim 12 \%$ y $\sim 8 \%$, respectivamente, para el modelo anual. Las mayores mejoras se observan en las regiones extra-tropicales (latitudes entre 30° y 55° N/S), donde las reducciones alcanzan aproximadamente el 16 % frente a ASCAT y entre el 11 % y el 12 % frente a HSCAT, lo que refleja estructuras de sesgo más predecibles en estas regiones. En los trópicos (latitudes $<30^\circ$), los modelos mensuales muestran una reducción de la varianza del error de alrededor del 10 % validando con ASCAT, aunque las mejoras son más modestas al evaluarse frente a HSCAT-B. En altas latitudes ($>55^\circ$ N/S), el rendimiento es similar para ambos tipos de dispersómetro, con reducciones de la varianza del error en torno al 10–11 %.

La Figura 2 muestra la distribución espacial de la reducción/incremento del VRMS de la red neuronal mensual en comparación con ERA5, cuando se validan

Tabla 2. Reducción de la varianza del error (%) para los modelos anuales y mensuales durante el período de prueba de 2023.

Red. var. error	Global		Trópicos		Extra-Trópicos		Latitudes Altas	
	Anual	Mensual	Anual	Mensual	Anual	Mensual	Anual	Mensual
ASCATB	12.19	12.97	10.24	11.49	15.44	15.83	10.39	10.88
ASCATC	12.44	13.22	10.48	11.82	15.77	16.07	10.51	10.95
HSCATB	7.88	8.60	3.81	4.90	11.71	12.03	9.35	10.00
HSCATC	8.77	9.05	5.10	6.09	11.29	11.22	10.60	10.29

frente a ASCAT-B y HSCAT-B. Las áreas amarillas indican un impacto neutro, mientras que las regiones verdes (rojas) señalan zonas donde el rendimiento mejora (empeora) con respecto a ERA5. La validación frente a dispersómetros en banda C revela mejoras notables en torno a la ZCIT. En las regiones extra-tropicales, las mejoras más significativas se observan en las partes occidentales de las cuencas oceánicas del hemisferio norte, correspondientes a regiones con fuertes corrientes superficiales zonales. En el hemisferio sur, la mejora es espacialmente más homogénea. La validación frente a HSCAT-B, sin embargo, muestra mejoras más modestas y cierta degradación en los trópicos, principalmente en la región de la ZCIT.

Estas diferencias se deben principalmente a los distintos tiempos locales de muestreo de los instrumentos: ASCAT observa en torno a las 09:30 hora solar local, mientras que HSCAT-B muestrea alrededor de las 06:00. Dado que ERA5 presenta errores asociados al ciclo diario, la red neuronal entrenada exclusivamente en los horarios de ASCAT captura y corrige mejor las estructuras de sesgo presentes en estas horas. Otras diferencias se derivan de la menor resolución espacial y la mayor sensibilidad a la precipitación de los sistemas en banda Ku, que modifican el muestreo meteorológico efectivo tras la aplicación del control de calidad.

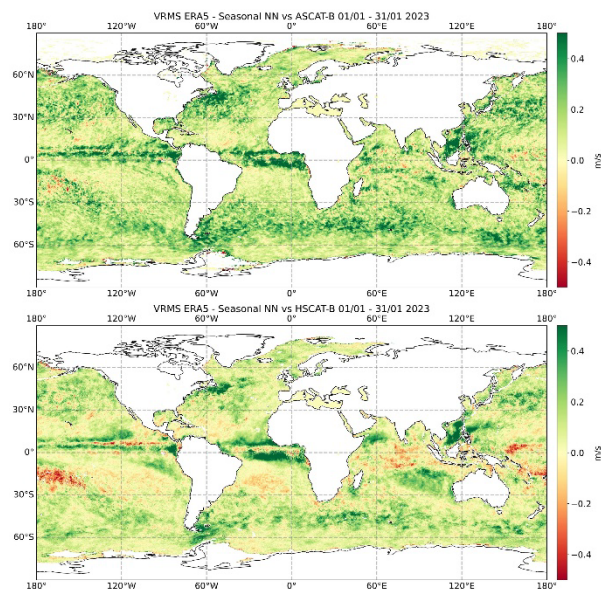


Figura 2. Reducción media (verde) o incremento (rojo) del VRMS de la salida de la red neuronal mensual (en comparación con ERA5), verificada frente a ASCAT-B (arriba) y HSCAT-B (abajo).

4. CONCLUSIONES

Este estudio demuestra que los modelos de aprendizaje automático pueden reducir eficazmente los sesgos persistentes de los vientos equivalentes al esfuerzo en ERA5 sin requerir observaciones de dispersómetros durante la inferencia. Al entrenar las FCNN con cinco años de datos atmosféricos y oceánicos, los modelos capturaron relaciones sistemáticas entre las variables de NWP y los patrones de sesgo del viento con mejoras significativas respecto al caso de referencia ERA5.

Los modelos mensuales proporcionan las mejoras consistentes, reflejando la variabilidad estacional de los patrones de lo sesgo. Las diferencias en la validación frente a ASCAT-B y HSCAT-B indican la influencia de las características de muestreo de cada instrumento, incluyendo sus distintos tiempos locales de observación y geometrías de visión, que explican parcialmente las mejoras más modestas obtenidas con los sistemas en banda Ku.

En trabajos futuros se explorarán arquitecturas más avanzadas, como las redes neuronales de grafos (GNN) (Lam *et al.*, 2023), funciones de pérdida mejoradas y la adaptación de los modelos de red neuronal a los datos del modelo operacional del ECMWF para la corrección de sesgos en los pronósticos futuros.

5. AGRADECIMIENTOS

Este estudio fue financiado por Organización Europea para la Explotación de Satélites Meteorológicos (EUMETSAT) Ocean Sea and Ice (OSI) Satellite Application Facility (SAF) Visiting Scientist Activity (OSI_VSA24_01).

6. REFERENCIAS

- Belmonte Rivas, M., & Stoffelen, A. (2019). Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT. *Ocean Science*, 15(3), 831–852. <https://doi.org/10.5194/os-15-831-2019>
- CMEMS. (2023). Global Ocean Hourly Sea Surface Wind and Stress from Scatterometer and Model. <https://doi.org/10.48670/moi-00305>
- CMEMS. (2023). Global Total (COPERNICUS-GLOBCURRENT), Ekman and Geostrophic currents at the Surface and 15m., <https://doi.org/10.48670/mds-00327>.
- Hersbach, H., *et al.* (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>
- Lam, L., *et al.* (2023), Learning skillful medium-range global weather forecasting. *Science*, vol. 382, no. 6677, pp. 1416–1421, <https://doi.org/10.1126/science.adi2336>.
- Makarova, E., Portabella, M., & Stoffelen, A. (2025). Reduction of persistent stress-equivalent wind biases with machine learning and scatterometer data. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–11. <https://doi.org/10.1109/TGRS.2025.3586375>
- Sandu, I., *et al.* (2020). On the causes of systematic forecast biases in near-surface wind direction over the oceans. ECMWF Technical Memorandum, 866.
- Trindade, A., Portabella, M., Stoffelen, A., Lin, W., & Verhoef, A. (2020). ERAStar: A High-Resolution Ocean Forcing Product. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2), 1337–1347. <https://doi.org/10.1109/TGRS.2019.2946019>