Response to Referee #1

Blue font indicates the reviewer's comments, black font represents responses to these comments, and red font marks the corresponding modifications in the manuscript.

The authors used a random forest and SHAP algorithms reconstructed a monthly sea surface wind in 1°×1° grids between 60°S-60°N from 1950-2023 based on ICOADS observations. The results were validated by ERA5, JRA-55, NOAA-20C, NCEP1, NCEP2, and Mn/Ca records in climatology, annual cycle, interannual variability, and long-term trend. The paper is well written and the dataset should be a good reference to user communities, and can be published in ESSD after a major revision. My major concerns are (a) lack of clarification on the use of CCMP as label variable, (b) lack of inter-comparisons against independent observations, and (c) lack of validation on wind direction.

Thank you for your valuable and constructive comments. Following your suggestions, the revised manuscript has been improved in the following aspects: (1) a detailed clarification on the use of CCMP as label variable, along with an evaluation of its consistency with in-situ observations, has been added (Table R1); (2) the Global Tropical Moored Buoy Array (GTMBA) has been adopted as an independent observation to evaluate the accuracy of the reconstructed data (Figs. R4-6); and (3) a comprehensive explanation of wind direction reconstruction and a validation on wind direction over the global oceans have been supplemented (Fig. R3). Our point-by-point replies to the specific suggestions are as follows.

1. L28-30, it is more important about the evaluation during the non-training period, as stated later in L32-35 the results are comparable. Therefore it should be clearly stated about the improvement of the current study.

Following your suggestion, the description of the evaluation results has been refined to highlight the improvement in the long-term trend (L34-L39 in the revised manuscript). Specifically, the long-term trends of MLAWind from 1950–1982 are corroborated by independent proxy indicators, exhibiting higher temporal correlation than those derived from existing reanalysis datasets. MLAWind offers a more reliable reconstruction of long-term sea surface wind variability during non-satellite periods.

“Independent coral records from 1950–1982 are employed as proxy indicators to evaluate long-term sea surface wind trends. The trend derived from MLAWind is corroborated by these proxies, exhibiting significantly higher temporal correlation than those obtained from reanalysis products. MLAWind demonstrates a more reliable reconstruction of long-term sea surface wind variability during historical periods since 1950.”

2. L90, “It” is not clear, is it “WASWind”?

We suppose that you are referring to the "it" in L96: "It exhibits a relatively coarse spatial resolution of 4°×4° and limited temporal coverage from 1950 to 2008.". Here, "it" refers to WASWind. In the revised manuscript, "it" has been replaced with "WASWind" to make the sentence in L94-L98 clearer:

“The Wave- and Anemometer-Based Sea Surface Wind (WASWind) achieves high accuracy by integrating anemometer-measured winds and wave height-estimated winds from the ICOADS data (Tokinaga and Xie, 2011). WASWind exhibits a relatively coarse spatial resolution of 4°×4° and limited temporal coverage from 1950 to 2008.”

3. L121, “an interpretable machine learning algorithm” need a reference.

An interpretable machine learning algorithm integrating the random forest model with SHapley Additive exPlanations (SHAP) interpretability module is employed in this study. The relevant references are added to the revised manuscript in L123-L124 and L762-L765:

“This study employs an interpretable machine learning algorithm (Tian et al., 2022; Kim et al., 2025) to integrate ICOADS and satellite data.”

“Kim, S., Alizamir, M., Heddam, S., Chang, S. W., Chung, I. M., Kisi, O., and Kulls, C.: Development of the machine learning and deep learning models with SHAP strategy for predicting groundwater levels in South Korea, Sci. Rep., 15, 35523, https://doi.org/ 10.1038/s41598-025-19545-y, 2025."

4. L136, WD, it is not clear why U, V, and WS were used in validation data while the inputs from ICOADS used WS and WD.

We sincerely apologize for any ambiguity that has arisen from the inadequate descriptions in the original manuscript. To address your questions on wind direction (WD), please find the detailed explanation below:

(1) The original ICOADS data provides observed wind speed (WS) and WD, yet it does not comprise the zonal (U) and meridional (V) wind components. **ICOADS WD was incorporated as an input feature, rather than serving as the reconstruction target.** Given that machine learning models do not necessitate parity between inputs and targets (Tian et al., 2022; Zhang et al., 2022), **U, V, and WS were reconstructed in this study**. These variables are routinely available from widely-used satellite and reanalysis products, enabling direct comparison between our results and these datasets in subsequent validation.

(2) Since WD can be derived from U and V (Eqs. R1-R3; Freeman et al., 2017), U and V inherently contain wind direction information. An independent reconstruction of WD is unnecessary.

$$WD = 180 + arctan2(U, V) \times \frac{180}{\pi} \tag{R1}$$

$$U = -WS \times sin\left(WD \times \frac{\pi}{180}\right) \tag{R2}$$

$$V = -WS \times cos\left(WD \times \frac{\pi}{180}\right) \tag{R3}$$

The relevant descriptions are refined to the revised manuscript in L137–L142 and L235–L238:

"The incorporation of physically relevant feature inputs significantly enhances the reconstruction accuracy of machine learning models (Tian et al., 2022; Zhang et al., 2022). Multiple variables closely associated with sea surface winds are utilized in this study, including sea surface wind speed (WS), wind direction (WD), SST, air temperature (Ta), and sea level pressure (SLP) from 1950 to 2023."

"Given that 10-m zonal wind, 10-m meridional wind and WS are routinely available from widely-used satellite and reanalysis products, these variables are selected as reconstruction targets to enable systematic inter-product comparison."

5. L156, "Smith (1980)" is messing in references.

The reference is added to the revised manuscript in L827-L829:

"Smith, S. D.: Wind stress and heat flux over the ocean in gale force winds, J. Phys. Oceanogr., 10, 709–726, https://doi.org/10.1175/1520-0485(1980)0102.0.CO;2, 1980."

6. L160-171, Have these satellite-based wind gone through a "bias-adjustment" process as for the in situ observations? How do we know these satellite observations are consistent with in situ observations?

The CCMP achieves bias adjustment of sea surface wind through systematic pre-processing and quality control procedures. Specifically, all satellite data integrated into the CCMP are consistently converted to 10-m neutral stability winds, followed by inter-satellite cross-calibration to minimize systematic biases and enhance the overall accuracy of the CCMP (Mears et al., 2019). Satellite sensors are calibrated using in-situ observations such as moored buoys. The observations are processed to correct for stability effects, and subsequently adjust the CCMP background surface wind field during the data assimilation phase (Mears et al., 2019).

The consistency between the CCMP and in-situ observations is evaluated based on the Global Tropical Moored Buoy Array (GTMBA). The GTMBA is a real-time, in-situ observational system that delivers reliable oceanic and atmospheric measurements across the tropical oceans (McPhaden et al., 2010). The CCMP shows strong agreement with the GTMBA from 1993–2023, exhibiting low mean bias and RMSE across all three wind components (U, V and WS). The Pearson correlation coefficients (CORR) exceed 0.98 for each variable (Table R1), further demonstrating the reliability of the CCMP satellite observations.

Table R1. Comparison of the CCMP satellite data and the GTMBA in-situ observations from 1993–2023. Mean bias, RMSE and CORR are used as quantitative metrics.

	U	V	WS
Bias (m s^{-1})	0.32	0.14	0.43
RMSE (m s^{-1})	0.40	0.39	0.42
CORR	0.99	0.99	0.98

More detailed descriptions of the CCMP are added to the revised manuscript in L178-L183:

"All satellite data are consistently converted to10-m neutral stability winds, followed by inter-satellite cross-calibration to minimize systematic biases. Moored buoys are used to calibrate the satellite sensors, adjusting the CCMP background surface wind field during the data assimilation phase. Therefore, the CCMP demonstrates strong agreement with in-situ observations (Mears et al., 2019)."

7. L208-211, I am not clear how the CCMP (1993-2023) is used as a label variable while ICOADS (1950-2023) used as feature input. What is the label variable during 1950-1992?

We sincerely apologize for any misunderstanding caused by the unclear description of the label variable and feature input. In this study, model training is conducted over the period 1993–2023, wherein the CCMP data from 1993–2023 serves as the label variable and the ICOADS data from 1993–2023 serves as the feature input. The nonlinear relationship between the CCMP winds and ICOADS variables is captured during model training.

Model inference is conducted over the period 1950–1992, during which the CCMP data are unavailable. The inference process operates without label variable. Leveraging the established relationship between the CCMP and ICOADS, sea surface winds during the non-satellite periods are reconstructed from ICOADS variables spanning 1950–1992.

In the revised manuscript, the descriptions are refined in L238-L243 and L275-L281 to clarify the definition of the feature inputs and label variables:

"Multiple ICOADS variables (WS, WD, SST, Ta, SLP) from 1993–2022 are used as feature inputs, while the CCMP 10-m wind field from 1993–2022 serves as the label variable for data reconstruction model establishment. The ICOADS data from 1950–1992 are employed in the model inference process to enable reconstruction of sea surface winds during historical periods without satellite data."

"The final step is the model inference process that generates the restructured dataset. The process operates without label variable. Following the training process conducted from 1993–2022, the machine learning model derives a nonlinear relationship between ICOADS variables and CCMP sea surface wind. The established relationship enables the reconstruction of historical sea surface winds during the non-satellite periods by utilizing ICOADS data from 1950–1992."

8. Figure 3a,c, I assume the color shading represent wind speed, my question is: what are the wind directions. Is there any metrics to measure the successes of the reconstruction? Clear differences can be found in wind speed in 2000 in the northern North Pacific and northern North Atlantic.

The color shading in Fig. 3 represents wind speed, and we apologize for the omission of wind direction annotations in Fig. 3a and 3c. Following your suggestion, wind direction (vector) and wind speed (shading) of the original ICOADS and MLAWind dataset are both presented in the revised manuscript (Fig. R1). The U and V components from the ICOADS are calculated based on WS and WD (Eqs. R2-R3). Sea surface winds derived from the ICOADS exhibit pronounced spatial sampling gaps, particularly in the Southern Hemisphere westerly belt. The MLAWind demonstrates continuous spatial coverage and smoother sea surface wind characteristics. It exhibits significant discrepancies relative to the ICOADS across the North Pacific and North Atlantic in 2000 (Fig. R1).

To measure the successes of the reconstruction, the CCMP satellite observation is employed to evaluate the biases of the ICOADS and MLAWind in 2000 (Fig. R2). The ICOADS exhibits significant systematic biases, with pronounced positive biases in the North Pacific and North Atlantic. The biases originate from uncertainties in ship height measurements and instrumental calibration protocols (Berry and Kent 2011; Tokinaga and Xie 2011). The MLAWind effectively mitigates these biases and demonstrates high consistency with the CCMP globally, confirming the validity of the reconstruction.

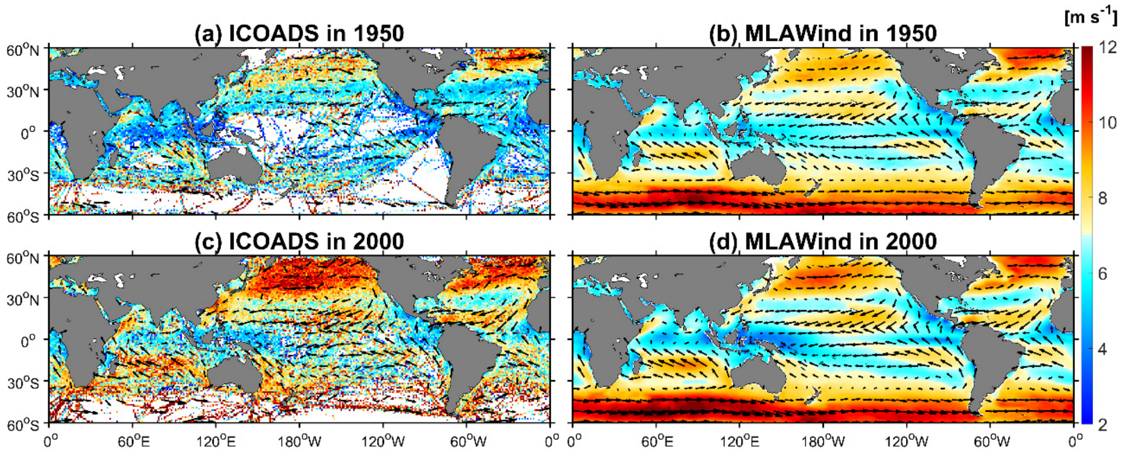

Figure R1. Reconstruction results presentation. 10-m wind velocity (vector, m s^{-1}) and sea surface wind speed (shading, m s^{-1}) of the original ICOADS and MLAWind dataset are compared in 1950 and 2000, respectively. These two years are chosen as examples to represent periods with sparse and abundant original data coverage, respectively.

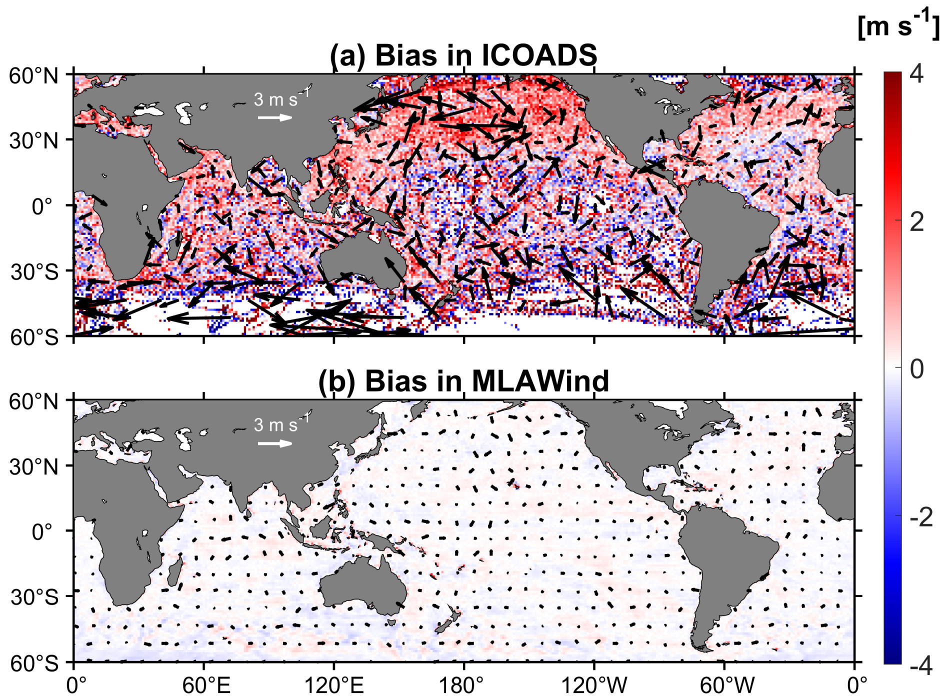

Figure R2. Bias of (a) the original ICOADS and (b) reconstructed MLAWind relative to the CCMP satellite observation in 2000. Vectors and shading in (a)-(b) indicate biases of 10-m wind velocity and sea surface wind speed, respectively.

Following your suggestions, wind direction represented by U and V vectors are added in Fig. 3a, c in the revised manuscript. The validation of the reconstruction efficacy is described in L292-L298:

"The CCMP satellite observation is used to evaluate the biases of ICOADS and MLAWind in 2000. ICOADS exhibits significant systematic biases, primarily attributed to uncertainties in ship height measurements and instrumental calibration protocols (Berry and Kent, 2011; Tokinaga and Xie, 2011). MLAWind effectively mitigates these biases and demonstrates high consistency with the CCMP, confirming the validity of the reconstruction."

9. L309-316, I suggest adding a table to compare these RMSE, R-squared, bias etc so that readers can easily understand the results.

Following your suggestion, a table is added to the revised manuscript in L372-L373 to compare R^2, mean bias, RMSE and sample size (N) on the training set and testing set:

Table 2. Comparison of MLAWind's performance on the training set (1993–2022) and testing set (2023). The CCMP satellite data is used as the evaluation benchmark.

Evaluation metric	Training set	Testing set
R^2	0.98	0.94
Bias (m s^{-1})	-0.01	0.18
RMSE (m s^{-1})	0.34	0.45
N	11257560	375252

10. Figure 4, is it possible to check the wind direction v/u as a verification metrics in Eqs (1)-(5)?

Wind direction (WD) can be evaluated using two metrics: bias and RMSE. Following your suggestion, a global evaluation of WD is performed from 1993–2022. The CCMP satellite data is employed to assess the errors of multiple datasets. MLAWind demonstrates the lowest mean bias across all evaluated datasets, with a value of only -0.1°. Its RMSE is slightly higher than that of ERA5, but remains lower than those of other reanalysis datasets. Despite the absence of direct reconstruction of WD, MLAWind derives WD with high accuracy (Fig. R3).

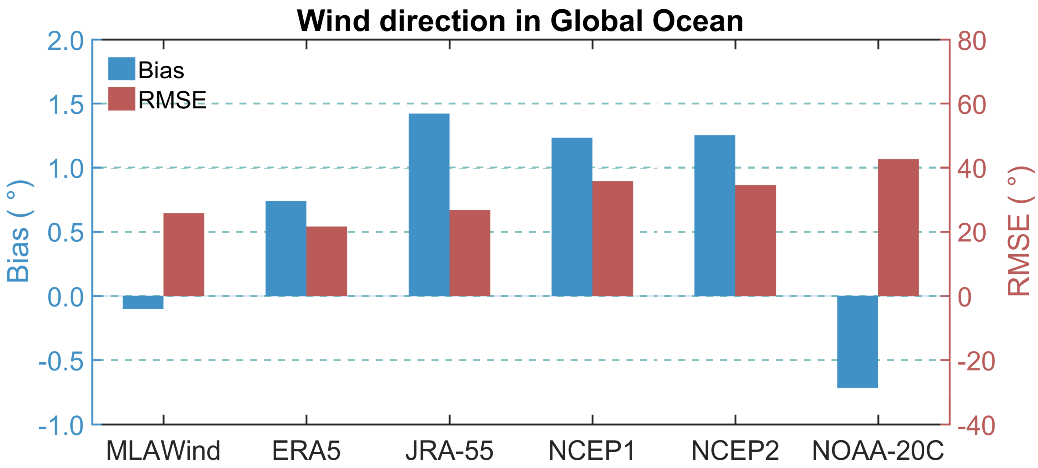

Figure R3. Evaluation of wind direction over the global oceans within 60°S–60°N from 1993–2022. The wind directions for all datasets are calculated based on Eq. R1.

The description of wind direction evaluation is added to the revised manuscript in L332-L337:

"A global evaluation of WD is performed to further confirm the reliability of the MLAWind dataset. The MLAWind dataset demonstrates the lowest mean bias and the second-lowest RMSE across all evaluated products. Despite the absence of direct reconstruction of WD, the MLAWind dataset derives WD with high accuracy."

11. Figures 7-10 are good for the intercomparisons. However, I am wondering the direct comparisons against observations particularly independent observations to see whether MLWind's performance is comparable with those reference datasets?

Following your suggestion, the Global Tropical Moored Buoy Array (GTMBA) is employed to assess the performance of MLAWind. The GTMBA is a real-time, in-situ observational system that delivers reliable oceanic and atmospheric measurements across the tropical oceans. It integrates three buoy arrays: the Tropical Atmosphere Ocean/Triangle Trans-Ocean Buoy Network (TAO/TRITON) in the Pacific Ocean, the Prediction and Research Moored Array in the Tropical Atlantic (PIRATA), and the Research Moored Array for African-Asian-Australian Monsoon Analysis and

Prediction (RAMA) in the Indian Ocean (McPhaden et al., 2010). These arrays exhibit distinct deployment periods and spatial coverage, yet are unified under common technical standards and data protocols. TAO provides a long-term observational record dating back to the early 1990s, whereas PIRATA and RAMA were deployed in 1997 and 2002, respectively. 56 TAO stations, 19 PIRATA stations, and 28 RAMA stations are utilized in this study (Fig. R4).

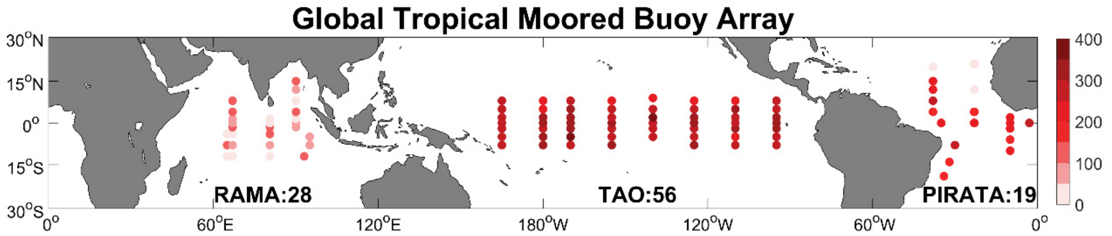

Figure R4. Distribution of the Global Tropical Moored Buoy Array (GTMBA). Monthly mean data are used, and the shadings at each station indicate the data volumes from 1993 to 2023.

The sea surface wind speeds derived from MLAWind exhibit strong agreement with in-situ observations from the RAMA, TAO, and PIRATA tropical buoy arrays, with all corresponding CORRs exceeding 0.90. Its performance is comparable across the Indian, Pacific, and Atlantic Oceans, characterized by a systematic bias of approximately 0.6 m/s and a RMSE below 0.45 m/s (Fig. R5). An evaluation of multiple datasets is conducted based on the GTMBA (Fig. R6). MLAWind and ERA5 exhibit smaller errors compared to other datasets. These two datasets equally achieve low RMSE (~0.40 m/s) and high CORR (0.95), although MLAWind demonstrates a slightly larger bias than ERA5. Note that ERA5 assimilates the GTMBA and thus its evaluation results are not independent. MLAWind achieves reconstruction accuracy comparable to ERA5, without relying on in-situ buoy observations.

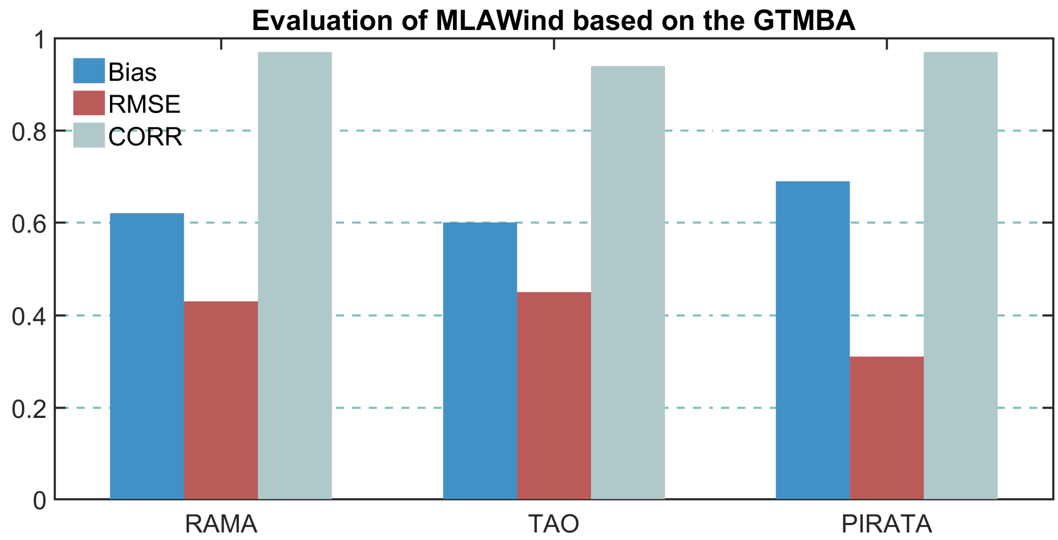

Figure R5. Evaluation of MLAWind based on the GTMBA observations. The evaluation periods for RAMA, TAO, and PIRATA are 2002–2023, 1993–2023, and 1997–2023, respectively.

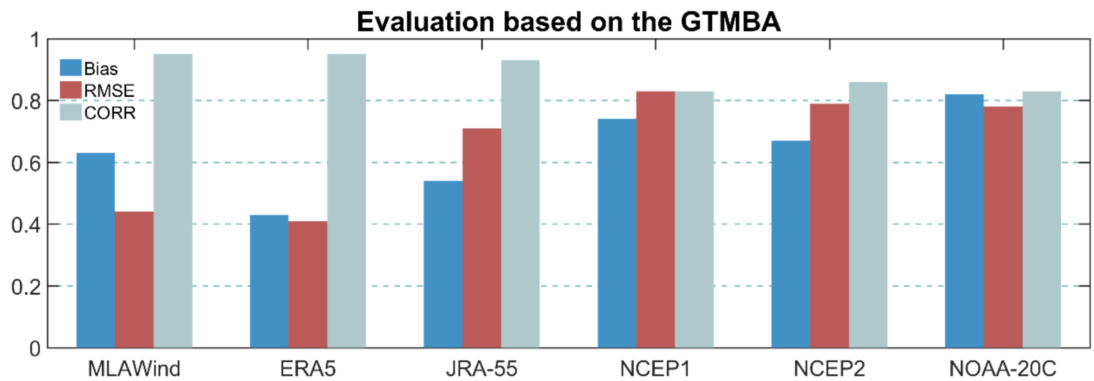

Figure R6. Evaluation of multiple datasets based on the GTMBA observations from 1993–2023.

The introduction and evaluation of GTMBA are added to the revised manuscript in L195-L206 and L374-L382, respectively:

"The Global Tropical Moored Buoy Array (GTMBA) in-situ observation is employed to evaluate the performance of multiple sea surface wind datasets. The GTMBA provides real-time observations in the tropics for weather and climate research. It integrates three buoy arrays, including the Tropical Atmosphere Ocean/Triangle Trans-Ocean Buoy Network (TAO/TRITON) in the Pacific Ocean, the Prediction and Research Moored Array in the Tropical Atlantic (PIRATA), and the Research Moored

Array for African-Asian-Australian Monsoon Analysis and Prediction (RAMA) in the Indian Ocean (McPhaden et al., 2010). TAO provides a long-term observational record dating back to the early 1990s, while PIRATA and RAMA were deployed in 1997 and 2002, respectively. 56 TAO stations, 19 PIRATA stations, and 28 RAMA stations are utilized in this study."

"In addition to satellite data, the performance of multiple datasets from 1993 to 2023 is compared based on the GTMBA in-situ observations (Fig. 7). MLAWind and ERA5 demonstrate high consistency with the GTMBA, exhibiting smaller errors than other datasets. These two datasets equally achieve low RMSE (~0.40 m/s) and high CORR (0.95), although MLAWind demonstrates a slightly larger bias than ERA5. Note that ERA5 assimilates the GTMBA and thus its evaluation results are not independent. MLAWind achieves reconstruction accuracy comparable to ERA5, without relying on in-situ buoy observations."

12. Table 2, I have difficulty to understand how the Mn/Ca trend can be quantitatively compared with wind because they are different variables in different units. Further, the difference among MLWind, ERA5, NCEP1, and NOAA-20C are large, why?

We sincerely apologize for any misunderstanding caused by the unclear description. Mn/Ca ratio in coral skeletons cannot be quantitatively compared with sea surface wind. However, it serves as a reliable qualitative proxy for assessing sea surface wind variability, particularly during periods lacking reliable observations (Shen et al., 1992). Previous studies have revealed a robust positive correlation between Mn/Ca ratios and zonal wind anomalies, attributable to the dependence of Mn/Ca ratios on sea surface wind-driven mixing (Thompson et al., 2015; Sayani et al., 2021).

In this study, the increasing Mn/Ca ratio from 1950–1982 indicates a notable strengthening of the westerlies. MLAWind and ERA5 both reveal a significant intensification of zonal winds ($p < 0.05$), whereas the other datasets fail to accurately capture the long-term trends of sea surface wind (Table R2). Moreover, MLAWind achieves the highest temporal correlation (0.48) with the Mn/Ca proxy among all evaluated datasets, indicating that it offers a more reliable and physically consistent reconstruction of long-term sea surface wind variability during non-satellite periods (Table R2).

The four evaluated datasets (MLAWind, ERA5, NCEP1, and NOAA-20C) show significant differences in capturing the long-term trend of sea surface wind, primarily attributed to discrepancies in their data assimilation models and technical approaches. First, reanalysis products employ different assimilation frameworks such as four-dimensional variational assimilation (4D-Var), whereas MLAWind is derived through machine learning algorithm. Different reconstruction methods may introduce discrepancies in the spatial distribution and magnitude. Second, the observations assimilated in the reanalysis datasets exhibit substantial heterogeneity in both volume and type. Previous studies have indicated that these differences lead to considerable uncertainty in the long-term trend of sea surface winds (Zhang et al., 2023).

Table R2. Long-term trends of coral Mn/Ca and 10-m zonal wind surrounding Tarawa Atoll (4°S–6°N, 166°–176°E) from 1950–1982. Positive values of coral Mn/Ca indicate an intensification of westerly winds. The temporal correlation coefficients between Mn/Ca and multiple datasets are assessed. All datasets are smoothed using a 11-year running mean filter to remove the impact of high-frequency signals.

Variable		Long-term trend	*P* value	Temporal correlation coefficient
Coral record (nmol mol^{-1}/decade)	**Mn/Ca**	**1.89**	***P*<0.05**	**-**
10-m U ($m\ s^{-1}$/decade)	**MLAWind**	**0.49**	***P*<0.05**	**0.48**
	ERA5	0.28	*P*<0.05	0.02
	NCEP1	0.14	*P*>0.05	-0.64
	NOAA-20C	0.14	*P*>0.05	-0.22

The descriptions of the relationship between Mn/Ca and sea surface wind are refined in L578-L592 in the revised manuscript:

"The pronounced increasing trend in Mn/Ca ratios from 1950 to 1982 suggests an intensification of the westerlies, yet it does not afford a quantitative estimate of the magnitude of sea surface wind variability. Significant uncertainties remain in the long-term trends across multiple datasets (Table 4), primarily attributed to discrepancies in their data assimilation models and technical approaches. Consistent with Mn/Ca records, the MLAWind and ERA5 datasets reveal a significant strengthening of westerly winds ($p < 0.05$), with a more pronounced trend observed in the MLAWind dataset. The trends derived from the NCEP1 and NOAA-20C datasets are weaker and statistically insignificant, demonstrating their relatively lower reliability for the period lacking satellite data. Moreover, the positive temporal correlation between MLAWind and Mn/Ca surrounding Tarawa Atoll is stronger than those observed in ERA5, NCEP1, and NOAA-20C, indicating that MLAWind provides a more reliable and physically consistent reconstruction of long-term sea surface wind variability during non-satellite periods (Table 4)."

References

Kim, S., Alizamir, M., Heddam, S., Chang, S. W., Chung, I. M., Kisi, O., and Kulls, C.: Development of the machine learning and deep learning models with SHAP strategy for predicting groundwater levels in South Korea, Sci. Rep., 15, 35523, https://doi.org/ 10.1038/s41598-025-19545-y, 2025.

McPhaden, M. J., Ando, K., Bourle`s, B., Freitag, H. P., Lumpkin, R., Masumoto, Y., Murty, V., Nobre, P., Ravichandran, M., Vialard, J., Vousden, D., and Yu, W.: The global tropical moored buoy array. In: Hall, J., Harrison, D. E., Stammer, D. (eds) Proceedings of the ‘‘OceanObs’09: Sustained Ocean Observations and Information for Society’’ conference, vol 2, Venice, Italy, 21–25 Sep 2009, ESA Publication WPP-306, https://doi.org/10.5270/OceanObs09.cwp.61, 2010.

Zhang, R., Guo, W., and Wang, X.: Deep Learning to Near-Surface Humidity Retrieval from Multi-Sensor Remote Sensing Data over the China Seas, Remote Sens., 14, 4353, https://doi.org/10.3390/rs14174353, 2022.