

Global Biogeochemical Cycles[®]

RESEARCH ARTICLE

10.1029/2024GB008344

Key Points:

- A high-resolution, spatio-temporally varying surface total alkalinity (TA) data product is constructed for the North Indian Ocean (NIO) (1993–2020)
- Deposition of acidic atmospheric dust in the northern part of the NIO may suppress the increase in surface TA
- Three sub-regions of the NIO exhibit significant increasing TA trends mainly driven by surface salinity trend

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. Chakraborty,
kunal.c@incois.gov.in

Citation:

Joshi, A. P., Ghoshal, P. K., Chakraborty, K., Roy, R., Jayaram, C., Sridevi, B., & Sarma, V. V. S. S. (2025). Long-term changes of surface total alkalinity and its driving mechanisms in the north Indian Ocean. *Global Biogeochemical Cycles*, 39, e2024GB008344. <https://doi.org/10.1029/2024GB008344>

Received 28 AUG 2024

Accepted 30 JUL 2025

Author Contributions:

Conceptualization: Kunal Chakraborty, Rajdeep Roy, V. V. S. S. Sarma

Data curation: Prasanna Kanti Ghoshal, Rajdeep Roy, Chiranjivi Jayaram, B. Sridevi

Formal analysis: A. P. Joshi, Prasanna Kanti Ghoshal

Investigation: A. P. Joshi, Prasanna Kanti Ghoshal

Methodology: A. P. Joshi, Kunal Chakraborty, V. V. S. S. Sarma

Resources: Kunal Chakraborty

Software: A. P. Joshi, Prasanna Kanti Ghoshal

Supervision: Kunal Chakraborty

Validation: Prasanna Kanti Ghoshal





Visualization: Kunal Chakraborty

Writing – original draft: A. P. Joshi

Writing – review & editing:

Kunal Chakraborty, Rajdeep Roy, Chiranjivi Jayaram, B. Sridevi, V. V. S. S. Sarma

Long-Term Changes of Surface Total Alkalinity and Its Driving Mechanisms in the North Indian Ocean

A. P. Joshi¹, Prasanna Kanti Ghoshal^{1,2}, Kunal Chakraborty¹ , Rajdeep Roy³ , Chiranjivi Jayaram³ , B. Sridevi^{1,4}, and V. V. S. S. Sarma⁴ 

¹Indian National Center for Ocean Information Services, Ministry of Earth Sciences, Hyderabad, India, ²KUFOS-INCOIS Joint Research Centre, Faculty of Ocean Science and Technology, Kerala University of Fisheries and Ocean Studies, Kochi, India, ³Regional Remote Sensing Centre-East, National Remote Sensing Centre, ISRO, Kolkata, India, ⁴CSIR-National Institute of Oceanography, Regional Center, Visakhapatnam, India

Abstract This study examines long-term changes in surface total alkalinity (TA) in the North Indian Ocean (NIO) by developing a machine learning-based data product (INCOIS_TA) using ship-based observations collected from different sources during the period 1978–2019 and a reanalysis data product. We identify three sub-regions within the NIO exhibiting significantly increasing TA trends, which are south of 7°N ($0.81 \pm 0.38 \mu\text{mol kg}^{-1} \text{ yr}^{-1}$), southeastern coast of the Arabian Sea (AS) ($1.16 \pm 0.42 \mu\text{mol kg}^{-1} \text{ yr}^{-1}$), and the southwestern region of the Bay of Bengal (BoB) ($0.47 \pm 0.26 \mu\text{mol kg}^{-1} \text{ yr}^{-1}$). The increasing trend in surface salinity primarily drives the rise in TA in these three regions. In other regions, upper ocean warming that increases stratification and weakens vertical mixing results in a reduced vertical supply of nutrients and high sub-surface TA, which may, in turn, lead to a decline in surface TA. However, this expected decrease is found to be insignificant, likely due to substantial atmospheric deposition of nutrients. Analysis of the annual mean excess alkalinity spatial pattern, after normalizing TA, indicates that the AS may be favorable for calcification, while the BoB may not. At interannual time scales, both the climatic modes (El Niño–Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD)) have significant and equal impact on the surface TA in the BoB, whereas in the AS, the changes in surface TA are dominated only by ENSO. Further, we find that strong El-Niño (La-Niña) and positive (negative) IOD years show a decrease (increase) in surface TA in the NIO.

Plain Language Summary This study develops a machine learning-based long-term (1993–2020), high-resolution gap-free data product of surface TA for the North Indian Ocean (INCOIS_TA). For the development of INCOIS_TA, we use ship-based observations, which are discretely available from 1978 to 2019. The study identifies three sub-regions (south of 7°N, the southwestern coast of the Arabian Sea (AS), and the southwestern region of the Bay of Bengal (BoB)) that show a significant increase in the TA trend. Regions apart from these sub-regions show a climate change-induced warming that increases stratification and weakens vertical mixing. Thus, we expect a declining TA trend. However, due to the atmospheric nutrient deposition, there is an insignificant change in productivity, which may result in an insignificant decrease in the TA trend. The removal of salinity from surface TA suggests the AS to be more conducive to calcification than the BoB. The ENSO primarily contributes to the interannual variability of surface TA in the AS, while both IOD and ENSO equally impact the interannual variability of surface TA in the BoB.

1. Introduction

The rise in atmospheric CO₂ due to anthropogenic perturbations has made a profound impact on marine ecosystems (Doney et al., 2020), making measurable changes in the ocean carbonate system (Le Quéré et al., 2010). This has led to an increase in hydrogen ion concentration ([H⁺]), as well as a decrease in carbonate ion concentration ([CO₃²⁻]) and in the saturation state of seawater (Ω) with respect to calcium carbonate minerals (CaCO₃) (Canadell et al., 2023; Le Quéré et al., 2009, 2010). The increase in hydrogen ion concentration is commonly described as a decrease in pH ($\text{pH} = -\log [\text{H}^+]$) and referred to as ocean acidification (Doney et al., 2009).

Total alkalinity (TA) is an essential variable in studying marine carbonate cycling in the oceans (Brewer et al., 1975; Feely et al., 2004; Sarmiento et al., 2002). It is defined as the number of moles of hydrogen ion equivalent to the excess of proton acceptors over proton donors in 1 kg of seawater (Dickson, 1992; Dickson et al., 2007; Dickson & Whitfield, 1981) and operationally defined by the titration of H⁺ of all weak bases present

in the solution (Dickson & Whitfield, 1981). The variability of TA in the surface ocean is primarily controlled by the addition of freshwater (e.g., river discharge, precipitation, and sea-ice melting) or removal (evaporation and sea-ice formation) (Brewer et al., 1986; Millero et al., 1998), which is also reflected in salinity distribution giving rise to unique salinity-alkalinity relationships in the global ocean (Fine et al., 2017; Krishna & Shanmugam, 2023; Lee et al., 2006; Metzl et al., 2024). Surface TA could be influenced by biological processes such as primary productivity and calcification (Barrett et al., 2025; Sarmiento & Gruber, 2006). Additionally, the biological formation and dissolution of calcium also play a role in modulating the TA, predominantly at greater depths (Lee et al., 2006; Wolf-Gladrow et al., 2007).

The North Indian Ocean (NIO) presents two distinct ecosystems and is strongly dominated by the reversal of monsoons, whereas this is less pronounced in the south (Vinayachandran et al., 2021). The presence of land-masses that limit the Indian Ocean (the Arabian Sea (AS) and the Bay of Bengal (BoB)) at relatively low latitudes ($\approx 25^\circ\text{N}$) significantly impacts its circulation pattern. The AS and the BoB also differ noticeably in terms of freshwater fluxes, as seen by surface salinity (S. P. Kumar & Prasad, 1999; Prasad & McClean, 2004). The AS experiences net evaporation, whereas net precipitation dominates in the BoB. In addition to that, BoB, including the Andaman Sea, gets the bulk of the runoff from the major rivers such as the Ganges, Brahmaputra, and Irrawaddy into the Indian Ocean, creating strong density stratification, which controls the carbonate dynamics of the BoB (Joshi et al., 2021; Joshi & Warrior, 2022; S. P. Kumar & Prasad, 1999; Prasad & McClean, 2004; Unesco, 1969).

In tropical and subtropical open oceans, changes in salinity account for more than 80% of the surface TA variability (Lee et al., 2006; Millero et al., 1998). The increase in atmospheric CO_2 increases surface ocean acidification, which impedes calcification and increases surface alkalinity (Barrett et al., 2025). At higher latitudes (i.e., north of 30°N or south of 30°S), a progressive increase in the convective mixing of deep waters rich in TA during seasonal cooling is an important additional factor that increases surface TA concentrations. The air-sea exchange of CO_2 does not affect alkalinity, as variability in ocean $p\text{CO}_2$ does not influence the TA of the ocean (Takahashi et al., 1981). The carbonate, bicarbonate, and borate ions in the surface ocean respond strongly to any changes in sea surface salinity (SSS). Therefore, TA and salinity remain well correlated, paving the way for developing empirical relations. Over the years, a number of attempts have been made to explore this TA and salinity relationship globally and regionally (Bates et al., 2006; Goyet et al., 1999; Krishna & Shanmugam, 2023; Lee et al., 2006; Lenton et al., 2012; Takahashi et al., 2014).

The availability of quality-controlled publicly accessible data sets makes it possible to attempt several new approaches (Bittig et al., 2018; Carter et al., 2017; Sauzède et al., 2017; Shaik et al., 2024) to improve the prediction of surface carbonate system variables (like TA and dissolved inorganic carbon (DIC)). These new approaches are being attempted based on essential ocean variables that are more often observed in the ocean, especially by autonomous floats (Metzl et al., 2024; Mignot et al., 2023). Mapping methods have thus become an essential tool in ocean carbon cycle research, allowing us to interpolate or extrapolate these limited measurements into space-time-varying fields of carbonate system variables. Despite these efforts, most of the global approaches show considerable biases in tropical and subtropical waters (Fine et al., 2017; Shaik et al., 2024), particularly in the NIO, necessitating the development of more basin-specific models.

The development of global surface TA maps was initialized with the works of Millero et al. (1998) and Lee et al. (2006). Millero et al. (1998) established a linear regression relation between salinity normalized TA and sea surface temperature (SST), which resulted in imprecise trends as noted by Friis et al. (2003). To capture the non-linear relationship between SST, SSS, and TA, Lee et al. (2006) introduced a quadratic relationship between these variables. This indeed improved the accuracy but failed to capture finer-scale variability near the coasts. Using data from the World Ocean Circulation Experiment (WOCE) and applying a linear regression relation between TA, SST, SSS, nitrate (NO_3), and oxygen (O_2), Bates et al. (2006) produced a climatological surface TA map for the Indian Ocean region. The TA and DIC estimates were subsequently used to calculate surface CO_2 fluxes. To include the effect of net productivity on surface TA, Takahashi et al. (2014) used potential alkalinity, that is, TA and nitrate combined, and formed a linear regression with SSS.

With increased ship-based observations of TA and its predictors (SST, SSS, and nutrients), many studies either directly used or updated the regression relationship and produced TA maps on regional and global scales (Carter et al., 2017, 2018; Goyet et al., 1999; Krishna & Shanmugam, 2023; Sasse et al., 2013). However, none of the studies could adequately produce the surface TA variability in two contrasting basins of the NIO. Chau

et al. (2024) used the locally interpolated alkalinity regression version 2 (LIARv2) approach developed by Carter et al. (2018) to predict the surface TA using SST, SSS, nitrate, and silicate as predictors. Carter et al. (2021) provides an updated version of LIAR, along with a neural network approach to determine carbonate variables, such as TA. However, the products developed using the approaches mentioned above have high uncertainties in the coastal regions due to the unavailability of a sufficient number of observations and the high variation in SSS. The linear regression equation often fails in coastal regions and low saline regions like the BoB (Takahashi et al., 2014), where the SSS variability is high due to the freshwater influx.

Machine learning (ML) algorithms have the ability to better capture nonlinear relationships. Hence, Broullón et al. (2019) developed a global TA climatology using a neural network approach. Gregor and Gruber (2021) used the Geospatial Random Cluster Ensemble Regression technique to form regional clusters over which regression equations are applied to predict the surface TA. Further, Shaik et al. (2024) employed the Tabular Network (TabNet) ML technique for estimating surface TA using satellite data. Mostly, the global products have a spatial resolution of $1^\circ \times 1^\circ$, except for Chau et al. (2024), which provides monthly surface TA at 0.25° spatial resolution. The inability of coarse resolution products to adequately incorporate changes in surface TA due to changes in local-to-regional scale processes, the failure of linear regression in high SSS variability regions, the paucity of near-coast observations, and discontinuous long-term availability of satellite data often restrict the researchers from having a comprehensive analysis of long-term changes in surface TA in a region like NIO.

In an effort to improve regional predictions of surface TA in the NIO and to study its long-term changes and variability across different timescales, we developed a high-resolution ($1/12^\circ$) surface TA product for the period 1993–2020 using an advanced ML algorithm. The availability of a significant number of coastal and open-ocean observations, combined with the use of a boosting ML technique, enabled the creation of a high spatial resolution product that captures region-specific TA variability driven by local and regional ocean processes. Since this TA product is based on observations, it includes both organic and inorganic components of alkalinity. As a result, using this TA to calculate other carbonate system variables may lead to more accurate predictions, particularly in coastal regions (Kerr et al., 2021). We compared this product with recently developed global ML-based data products to evaluate the improvements in our region-specific version. Further, we used this data product to investigate the seasonality and interannual variability of surface TA in the NIO.

2. Data and Methodology

2.1. Cruise Data Collection

Several global surveys in the Indian Ocean were conducted over the years, such as the WOCE, and Indian Joint Global Ocean Flux Study where measurements of marine inorganic carbonate system variables along with physical variables were carried out (Morrison et al., 1998; Sarma, 2002, 2006; Sarma et al., 1998, 2010; Sarma & Narvekar, 2001; Sridevi & Sarma, 2021). In addition to these observations, we collected observations of surface TA from several other scientific cruises that were undertaken in both AS and BoB. Detailed descriptions of in situ measurements and some of these data sets' calibration procedures are already published and detailed elsewhere (Goyet et al., 1999; Key et al., 2004; Millero et al., 1998; Sarma, 2003; Sarma et al., 2009).

Apart from these data sets, we also extracted surface TA data from the Global Ocean Data Analysis Project (GLODAP)v2.2022 (Key et al., 2015; Lauvset et al., 2022; Olsen et al., 2016). The TA data from GLODAP have been adjusted and compiled to ensure consistency within $\pm 4 \mu\text{mol kg}^{-1}$ (Lauvset et al., 2022). The accuracy of the TA measured from the Indian cruise lines is $\pm 5 \mu\text{mol kg}^{-1}$ (Sridevi & Sarma, 2024). Although only data from the upper 5 m were used in this study, the non-GLODAP data were collected throughout the entire water column. To assess the consistency between the GLODAP and non-GLODAP data sets, we compared deepwater measurements ($> 2000 \text{ m}$), assuming that TA remains stable at these depths over the timescale considered ($< 10 \text{ years}$). First, temperature and salinity data were examined to confirm that the same water mass was sampled in both data sets. Then, the TA data were compared. The differences between the GLODAP and non-GLODAP TA values were within the measurement errors, indicating that the two data sets are consistent with each other. In Figure 1, we demonstrate all the cruise lines that are used in this study, along which the observations were made. These cruise lines have decent spatial coverage. The addition of coastal observations from the Indian coasts enhances the spatial coverage of GLODAP alone. The GLODAP surface TA database (which is frequently used to develop global ML-based products (Chau et al., 2022, 2024; Gregor & Gruber, 2021)) does not include data along the

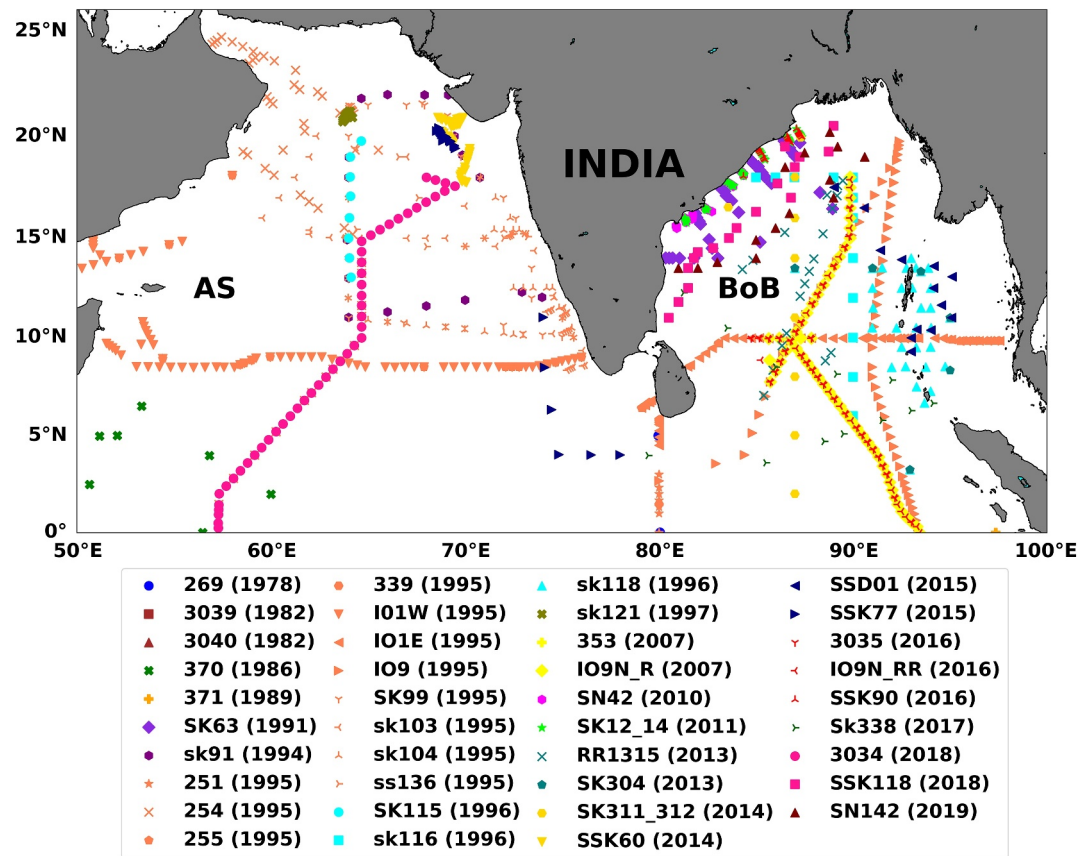


Figure 1. Map of the study region showing all the cruise lines used in this study. Each cruise line is represented by a different symbol. The color represents the year corresponding to the cruise commencement year. The names of each cruise line are mentioned in the lower panel. The details of the cruises are mentioned in Table S1 in Supporting Information S1.

Indian coast (see Table S1 in Supporting Information S1 for the source of each cruise data used in this study). Therefore, considering the increased spatial coverage of observations (than GLODAP alone) used in this study, we postulate to develop a TA product with better accuracy than other ML-based existing products.

The total number of ship-based observations available for this study is 1144. These 1144 observations are distributed over the whole NIO domain ($\approx 55.89\%$ in the BoB and $\approx 44.11\%$ in the AS) and span from 1978 to 2019. Although the data covers a long period, it is essential to understand the monthly distribution of this data (Figure 2) to have an understanding of the possible uncertainty of the model's monthly predictions. Figure 2 shows that April has the maximum number of observations, followed by February and September. November has the lowest number of observations, whereas no observations are available for December. Overall, the observations are well distributed during the pre-monsoon (March–May) and southwest monsoon (June–September) periods. However, the observations during the post-monsoon (October–November) and the northeast monsoon (December–February) are not well distributed.

2.2. Reanalysis Data

In the present study, we assume that the surface TA in the NIO is dependent on SSS, SST, and the mixed layer depth (MLD). These predictors serve as a proxy for temperature-induced biological effects, salinity-driven dilution, and mixing effects from the variations in the MLD on surface TA. Since the aim of this study is to produce gap-free long-term TA for NIO, we select our predictors (collocated at the cruise lines) from the GLORYS12V1 (Global Ocean Physics Reanalysis (GLORY)) product, provided by Copernicus Marine Services (CMEMS). The GLORY data from 1993 to 2021 has a horizontal resolution of 0.083° . Readers are encouraged to refer to Drévilion et al. (2022) for further details on this product.

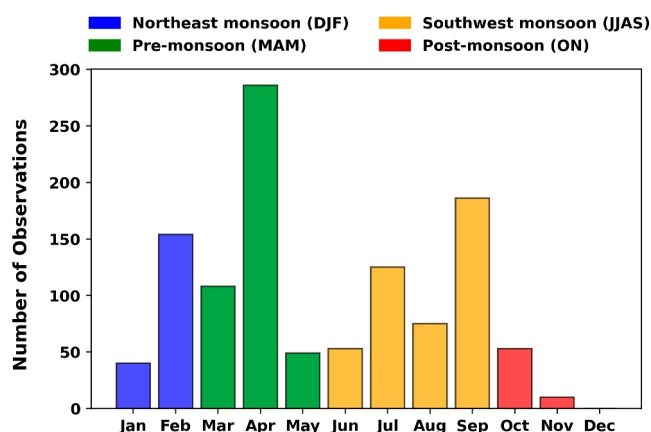


Figure 2. The monthly availability of surface total alkalinity observations in the North Indian Ocean. The blue color represents the northeast monsoon season (December–February (DJF)). Similarly, the green, yellow, and red colors represent the pre-monsoon (March–May (MAM)), southwest monsoon (June–September (JJAS)), and post-monsoon (October and November (ON)) seasons, respectively.

Nitrate (NO_3^-) is taken from the Global Ocean Biogeochemistry Hindcast provided by CMEMS (Perruche, 2018). It is an output of the coupled NEMO-PISCES simulations having a horizontal resolution of $1/4^\circ$. No data assimilation is done for this product. This product is available on a daily and monthly scale from 1993 to 2020. To maintain consistency, we use the data on a monthly scale for this study.

2.3. Methods

The complete methodology employed for this study is depicted using the flowchart (Figure 3). Each of the steps of the flowchart is described in this section.

2.3.1. Description of Different Algorithms or Routines

This section describes the algorithms or routines used in this study.

2.3.1.1. Extreme Gradient Boosting (XGB) Algorithm

The Extreme gradient boosting (XGB) algorithm (Chen & Guestrin, 2016) belongs to the family of decision tree-based boosting algorithms. Not only does the XGB algorithm improve predicting capabilities, but it also reduces computational efforts. It sequentially adds the weak learners by feeding the residuals to the next weak learner. This approach reduces the prediction error of the XGB algorithm. Instead of gradient descent, the algorithm uses Newton boosting based on the Newton-Raphson method, which reaches the global minima in an accelerated fashion. Further, the XGBoost includes advanced regularization schemes, which prevent overfitting. The advanced regularization scheme assists in handling custom loss functions and methodically handles the missing values. In our previous studies (Joshi et al., 2022, 2024), the XGB algorithm has outperformed other models when predicting ocean carbonate state variables. To avoid redundancy, we encourage readers to refer to Joshi et al. (2022) for the mathematical description of the XGB algorithm. This algorithm has some tunable hyper-parameters, which we have optimized (Table 1) using the Optuna optimization framework (Akiba et al., 2019). The total data is randomly split into an 80:20 ratio (i.e., 80% of the total data forms the training data set and 20% forms the test data set (Figure S1 in Supporting Information S1) using the sklearn library (Pedregosa et al., 2011). The test data is kept separately and used only to test the performance of the final prepared XGB model.

2.3.1.2. Empirical Seawater Property Estimator Routines

Trained on the data acquired from GLODPAv2.2020, the Empirical Seawater Property Estimator Routines (ESPERs) algorithm is able to predict up to seven biogeochemical ocean state variables such as DIC, TA, pH (Total Scale), phosphate, nitrate, silicate, and oxygen (Carter et al., 2021). The ESPERs provide two algorithms, the first one is based on linear regression (ESPER_LIR) and the second based on the neural network approach (ESPER_NN). The ESPERs is an upgrade on the previously known LIARv2 (Carter et al., 2018), which is used by Chau et al. (2024) to predict global surface TA at 0.25° spatial resolution. ESPERs provides 16 equations using a combination of 6 input ocean state variables such as temperature, salinity, oxygen, nitrate, silicate, and phosphate. For this study, the measurements of collocated temperature and salinity are provided from the GLORY reanalysis product (as mentioned in Section 2.2). Following Chau et al. (2024), the required nutrients are taken from World Ocean Atlas 2018 (WOA18). The ESPERs algorithm allows us to compare the performance of the XGB algorithm described in the previous section.

2.3.2. Ensemble Approach

Chau et al. (2022, 2024) introduced an ensemble approach to predict surface carbonate variables, similar to the Monte-Carlo statistical methods. The ensemble approach was recently used to correct the numerically modeled surface partial pressure of CO_2 in the Indian Ocean region (Ghoshal et al., 2025). This approach creates 150 randomly generated train-test (80:20) splits from the train set generated while optimizing the hyper-parameters. Then, using the tuned hyper-parameters of the XGB algorithm, we produced 150 models. We predict the surface TA of a common test set (generated earlier during the tuning of the hyper-parameters) using one model (out of the

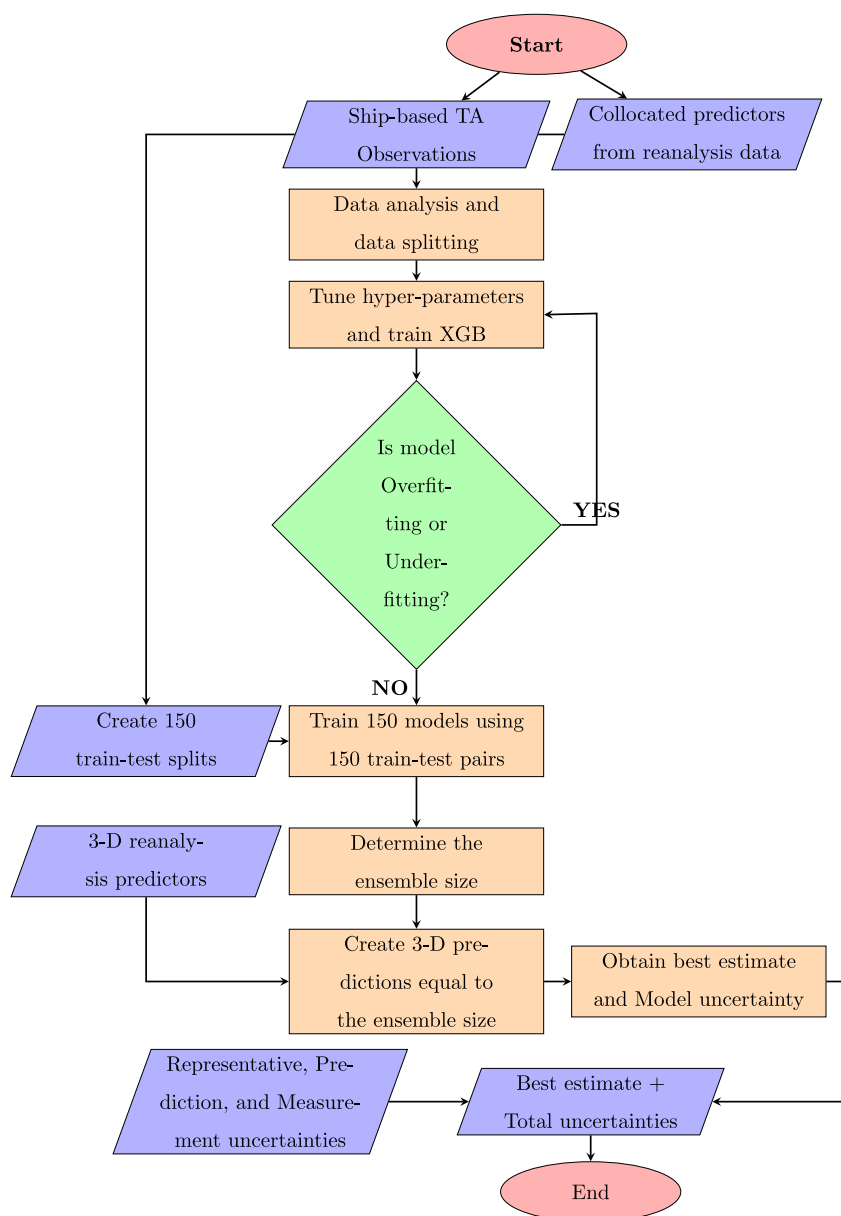


Figure 3. Chart showing the steps adopted in this study to prepare the final total alkalinity (TA) data product. The red ellipse indicates the beginning and end of the TA prediction process. Orange rectangles denote processes, blue trapeziums represent input/output, and green diamond-shaped boxes signify decisions made during the TA predictions.

150 models) at a time. We progressively average (ensemble) the prediction from the second model to the 150th model. We test the root mean square error (RMSE) between each ensemble size (e.g., if we ensemble the predictions from the first two models, then the ensemble size is 2. Similarly, the ensemble size increases as we add the predictions from other models) using the common test set of surface TA. The ensemble size after which the RMSE has negligible changes when compared with the RMSE obtained from the predictions of the previous size, is identified as the ideal ensemble size (Figure 4).

According to Goodhue et al. (2012), a minimum of 50 ensemble sizes is required to reach a stable cost function. As the size increases beyond 100, the difference between the cost function of the two ensemble predictions is negligible. This statistical information could be problem-specific, so we went up to 150 ensemble size. Figure 4 shows that the RMSE has been stabilized since 100 ensemble size. To save some computational effort, we select 140 as the ensemble size for this study. Further, as shown in Chau et al. (2022), the mean of the 140 predictions

Table 1
Range or Options and Final Optimized Values of the Extreme Gradient Boosting Hyper-Parameters

Hyper-parameters	Range or options	Optimized value
lambda	0–1	0.6
alpha	0–1	0.6
subsample	0–1	0.7
Booster	gbtree/gblinear/dart	gbtree
colsample_bytree	0–1	0.7
max_depth	10–500 (step = 1)	100
min_child_weight	1–100	25
learning_rate	1×10^{-2} –1	0.025
gamma	1×10^{-8} –1	0.01
n_estimators	100–800 (step = 1)	500
grow_policy	depthwise/lossguide	lossguide

will provide the best estimate, and the standard deviation of the ensemble predictions is known as the model uncertainty.

Thus, using the spatiotemporal predictors (SST, SSS, and MLD) from the reanalysis product, we predict the TA using each of these 140 models. This results in 140 estimates of monthly, $1/12^\circ$ surface ocean TA for the NIO region. Then we average all 140 spatiotemporal TA estimates to produce the best estimate of TA. The following equation describes this process of getting the best estimate of TA.

$$\mu_{tij} = \frac{\sum_{q=1}^{140} TA(t, i, j, q)}{140} \quad (1)$$

where μ_{tij} represents the best estimate of TA. The TA is spatiotemporally varying. Hence, t , i , and j represent the time, longitude, and latitude.

2.3.3. Uncertainty Calculation

Four sources of uncertainty are identified in this study. These sources are as follows:

1. Model Uncertainty (M)

Following Ghoshal et al. (2025) and Chau et al. (2024), we determine the model uncertainty. It is the standard deviation ($1-\sigma$) of the 140 spatiotemporal TA estimates. Mathematically, it is represented as follows:

$$M_{tij} = \sqrt{\frac{\sum_{q=1}^{140} [TA(t, i, j, q) - \mu_{tij}]^2}{140}} \quad (2)$$

where M_{tij} represents the spatiotemporal Model uncertainty.

2. Representation Uncertainty (R)

The ship-based measurements are coarse in space and time. The final predictions of TA will be monthly data (from 1993 to 2020) having $1/12^\circ$ spatial resolution. However, this resolution may be too coarse to accurately capture the typical scales of TA variability in the NIO region. This issue is particularly pronounced when the number of cruise observations within a grid cell is low, and spatio-temporal variability is high. Specifically, if cruise TA observations are temporally discrete, a grid cell for a given month and year may be represented by only one or a few days of sampling. Given our limited understanding of the spatio-temporal variability of TA from cruise data, precisely quantifying the representation error by directly comparing predicted and observed values is challenging. Therefore, following Gregor and Gruber (2021), we estimate the representative error of predicted TA using spatio-temporal gradient analysis.

To quantify this error, we change the spatial (using an interpolation method) and temporal resolution of the predicted surface TA from $1/12^\circ$ to $1/24^\circ$ and from monthly to 15 days, respectively. Then we calculate the spatial and temporal gradients across time and space, which are represented by the following equations.

$$\Delta_t(t, x, y) = \left| \frac{\partial TA}{\partial t} \right| \approx |TA(t + \delta t, x, y) - TA(t, x, y)| \quad (3)$$

$$\Delta_x(t, x, y) = \left| \frac{\partial TA}{\partial x} \right| \approx |TA(t, x + \delta x, y) - TA(t, x, y)| \quad (4)$$

$$\Delta_y(t, x, y) = \left| \frac{\partial TA}{\partial y} \right| \approx |TA(t, x, y + \delta y) - TA(t, x, y)| \quad (5)$$

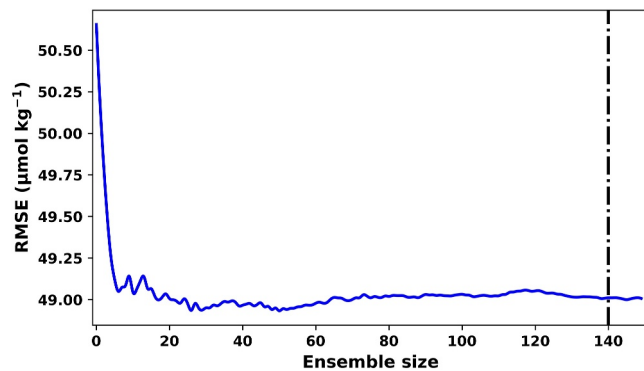


Figure 4. This figure depicts the changes in root mean square error (RMSE) with respect to the ensemble size. An increase in ensemble size stabilizes the RMSE. The vertical line denotes the ensemble size 140, chosen for this study.

where Δ_t , Δ_x , and Δ_y represent the gradients along time, longitude, and latitude, respectively.

The total spatial gradient (SG_s) is calculated as the mean of absolute differences along the longitude and latitude. Mathematically, it can be represented as follows.

$$SG_s(t, x, y) = \frac{1}{2}(\Delta_x(t, x, y) + \Delta_y(t, x, y)) \quad (6)$$

Then the representative error at fine grid scale ($1/24^\circ$, R_{fine}) is calculated as follows.

$$R_{\text{fine}}(t, x, y) = \frac{1}{2}(SG_s(t, x, y) + \Delta_t(t, x, y)) \quad (7)$$

Then, we re-grid R_{fine} back from higher resolution ($1/24^\circ$ and 15 days) to the original ($1/12^\circ$ and monthly) resolution to estimate representative error (R) (Gregor & Gruber, 2021).

Since the representative error is an error due to the local spatial and temporal scale. This error can be assumed to be fundamentally and mathematically separate from model uncertainty.

3. Measurement Uncertainty (MeU)

The instrumental error while making observations needs to be added to quantify total uncertainty. As mentioned in Section 2.1, $\pm 5 \mu\text{mol kg}^{-1}$ ($1-\sigma$) is the error in the TA measurements for the data collected from the Indian cruises, and $\pm 4 \mu\text{mol kg}^{-1}$ is the mean possible error from GLODAP TA measurements. In this study, we use $\pm 5 \mu\text{mol kg}^{-1}$ as the MeU. We assume that the error affects the whole domain uniformly and add this error to each grid cell. Since MeU comes from instrumental and/or analysis error, there is no overlap with the previous two uncertainties.

4. Input Uncertainty (I)

Since the best estimate of TA is produced using reanalysis data, it is critical to estimate the possible uncertainties in TA originating from the uncertainties in the reanalysis data. The error in the SST and SSS from the reanalysis data could be estimated from the collocated observation of these state variables. As shown in Figure 5, the RMSE for SST is 1.06°C and 1.58 for the SSS.

Since it is not possible to have such estimates of MLD, we follow Karmakar et al. (2018) approach to calculate the possible error in MLD. Karmakar et al. (2018) compares the MLD in the tropical Indian Ocean from various global reanalysis products to the MLD from the EN.4.2.0 data. For this study, we use the latest EN.4.2.2 data (Gouretski & Cheng, 2020) and calculate the RMSE with GLORYS12V1 MLD (Section 2.2) for the NIO region. The spatially averaged RMSE found for the NIO region is 12.22 m (Figure S2 in Supporting Information S1), which is then assumed as an error that may propagate in finding the best estimates of TA.

As mentioned in Section 2.3.2, we used 140 independent XGB models to create the best estimate of TA. To calculate the input uncertainties, we first created 100 perturbed realizations by adding a Gaussian noise (using the error of each input and a zero mean) to each of the input variables (following Wu et al. (2024)). Then, for each perturbed value of each input variable, we created an ensemble of 140 XGB models (same as the calculation of the best estimated TA (Equation 1)). Further, the difference between this ensemble and the best estimated TA was computed (Δ_i , where i is the input variable of interest). This process was repeated 100 times for each input variable ($100 \times \Delta_i$). Finally, the standard deviation of the $100 \times \Delta_i$ quantifies the uncertainty due to each of the input variables. The input uncertainty (I) is the quadrature sum of the uncertainty corresponding to each input variable (I_{SST} , I_{SSS} , and I_{MLD}), as represented by the following equation.

$$I = \sqrt{I_{\text{SST}}^2 + I_{\text{SSS}}^2 + I_{\text{MLD}}^2} \quad (8)$$

Both model and input uncertainties use an ensemble approach. However, model uncertainty comes from various train-test splits, but input uncertainty arises from the added perturbation using Gaussian noise. Representation error arises from the local spatio-temporal scale, while measurement error arises from the instrument and/or lab analysis. Hence, they can be assumed to be independent of the input uncertainty.

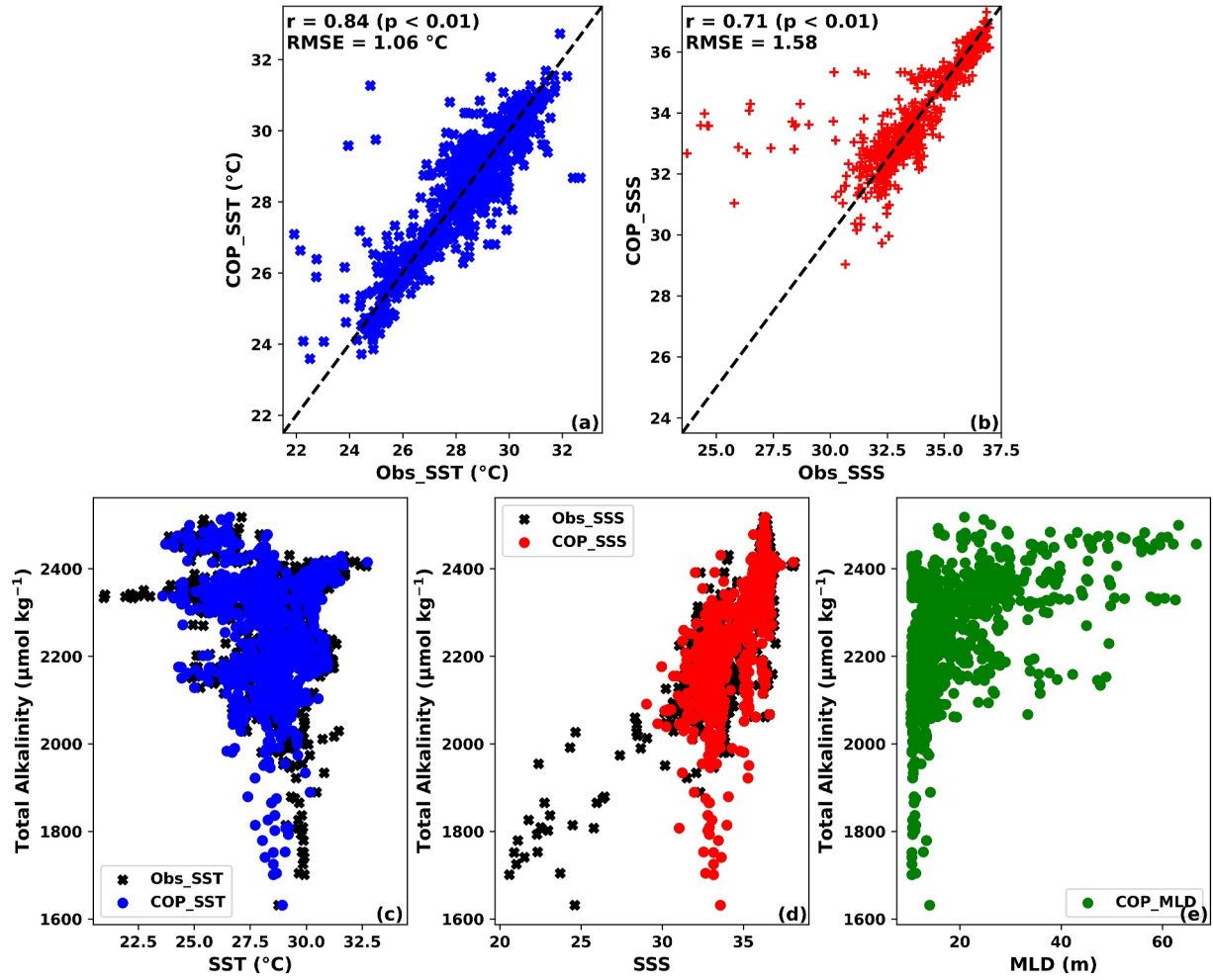


Figure 5. The first row of the figure shows a comparison between (a) ship-based sea surface temperature (SST) observations (Obs_SST) and reanalysis SST (COP_SST) and (b) ship-based sea surface salinity (SSS) observations (Obs_SSS) and reanalysis SSS (COP_SSS). The second row shows a relationship between surface total alkalinity of the North Indian Ocean with (c) reanalysis (COP_SST) and ship-based SST observations, (d) reanalysis SSS (COP_SSS) and ship-based SSS observations, and (e) only reanalysis mixed layer depth.

Thus, using the four identified sources of uncertainty and assuming they are independent of each other, we provide the total spatio-temporal uncertainty (ToU) as:

$$ToU = \sqrt{M^2 + R^2 + MeU^2 + I^2} \quad (9)$$

The readers should note that the best representation over the total period (T), that is, 1993 to 2020, of the spatio-temporal varying total uncertainty and its components is calculated using Equation 10.

$$U_{ij} = \sqrt{\frac{\sum_t U_{t,ij}^2}{T}} \quad (10)$$

where U represents total uncertainty or any of its components.

2.3.4. TA Trend With Corresponding Uncertainty

$$TA_{t,ij}^{(b)} \sim \mathcal{N}(\mu_{t,ij}, ToU_{t,ij}) \quad (11)$$

To estimate the mean TA trend and its associated uncertainty (due to the uncertainty in the TA predictions), we use a bootstrapping Monte-Carlo approach (as used while determining the model uncertainty). We generate synthetic TA time series ($TA_{t,i,j}^{(b)}$) at each grid cell by random sampling from a normal distribution. This normal distribution is defined by the best estimate of TA ($\mu_{t,i,j}$) and the associated total uncertainty (ToU) at each grid cell (Equation 11). A total of 10,000 bootstrapped samples were created. The number 10,000 was reached by progressively increasing the samples from 1,000 to 10,000. A further increase in sample size did not affect the mean TA trend (which should be the same as the trend of the best estimate of TA at each grid cell) and its associated uncertainty.

$$\beta^{(b)} = \frac{\sum_{i=1}^{10000} (t_i - \bar{t}) (T_i^{(b)} - \bar{T}^{(b)})}{\sum_{i=1}^{10000} (t_i - \bar{t})^2} \quad (12)$$

For each of the 10,000 synthetic TA data sets, we derive 10,000 mean-centered linear trends ($\beta^{(b)}$, see Equation 12) with respect to time (t , see Equation 12). The best estimate of the TA trend is defined as the average of these 10,000 linear trends, and the standard deviation provides the uncertainty associated with the best estimate of the TA trend.

2.3.5. Quantifying the Effect of the Trends of the Input Variables on the TA Trend

The changes in TA are primarily assumed to be driven by SST, SSS, and MLD in the NIO region. The effect of each of these variables can be analyzed using a decomposition analysis. The Taylor series expansion of this decomposition can be expressed as follows.

$$\frac{dTA}{dt} = \frac{\partial TA}{\partial SST} \frac{dSST}{dt} + \frac{\partial TA}{\partial SSS} \frac{dSSS}{dt} + \frac{\partial TA}{\partial MLD} \frac{dMLD}{dt} + nonlinear_effects \quad (13)$$

The left-hand side of Equation 13 represents the temporal variation of surface TA. The right-hand side of Equation 13 expresses the changes in TA due to the changes in each of the input variables. So, to quantify the effect of a particular input variable, we detrend the input variables except for the one whose effect we want to quantify. Then we reconstruct TA using the one variable of interest and the other two detrended input variables. The relative change between the trend of this reconstructed TA and the trend of the best estimated TA provides us with the required quantification (in percentage when multiplied by 100) of the effect of the input variable of our interest. For example, if we are investigating the effect of SSS trends on the TA trends, then we first detrend MLD and SST. Then we reconstruct TA (following the same process as determining the best estimated TA) using the SSS and detrended MLD and SST. Thus, the ratio of the trend of this reconstructed TA and the trend of the best estimated TA provides the effect of SSS trends on TA trends. This ratio, when multiplied by 100, will provide the contribution in terms of percentage.

2.3.6. Normalization of Surface TA

The TA in surface waters is controlled by evaporation, precipitation, river discharge, nutrient uptake, and remineralization (Fry et al., 2015). In this study, the TA data is normalized (to a constant SSS (35)) to remove the impact of all the above-mentioned processes following Fry et al. (2015) as shown in Equation 14.

$$nTA = \frac{(TA - TA_{river} + (1.36 \times [NO_3^-])) \times 35}{SSS} + TA_{river} \quad (14)$$

where Normalized TA (nTA) indicates normalized surface TA, TA_{river} is the mean river water TA, $[NO_3^-]$ is the Nitrate concentration in the NIO region. The impact of phytoplankton growth on TA is calculated by multiplying the $[NO_3^-]$ concentration by a factor of 1.36 (Fry et al., 2015).

The actual value of TA_{river} can be calculated as the intercept value of the $(TA + [NO_3^-])$ -SSS regression (see Equation 14). However, using such an intercept value may result in artificial variance or over-correction of TA due to the ignorance of non-dilution or non-concentration processes (caused by precipitation and evaporation).

(Friis et al., 2003; Jiang et al., 2014). Thus, salinity normalization of TA was proposed to be calculated based on a constant region-specific TA_{river} value (Friis et al., 2003).

Fry et al. (2015) used a value of $840 \mu\text{mol kg}^{-1}$ for TA_{river} to normalize TA in the Indian Ocean. They also tested a higher value of $1,106 \mu\text{mol kg}^{-1}$, based on measured TA from the Ganga-Brahmaputra outflow, but this led to an over-correction of TA in the northern BoB. Hence, an $840 \mu\text{mol kg}^{-1}$ value of TA_{river} was suggested to be used to normalize TA in the Indian Ocean region, and the same has been adopted for this study.

A limitation of normalizing TA is that it may not adequately remove the effect of the considered processes in all regions (e.g., regions having close to 35 PSU salinity). This normalization method considers that TA and SSS vary linearly, which may not be true in a real-world scenario. The use of a constant TA_{river} is an assumption, although it likely varies over time (e.g., seasonally) rather than across space. Since TA usually mixes conservatively, changes in TA near the coast are more likely due to biological processes rather than changes in river TA as you move away from the coast. However, we presume that this is the best method available in the literature to normalize the TA and proceed with the normalization using Equation 14.

After removing the effect of all the processes from TA, Fry et al. (2015) defines a tracer Alk^* to evaluate excess alkalinity patterns in the global surface ocean. The Alk^* is defined as follows.

$$Alk_{ij}^* = nTA_{ij} - \frac{\sum_{t,i,j} nTA_{ij} \times A_{ij}}{T \times \sum_{ij} A_{ij}} \quad (15)$$

where T represents total time length (months) and A_{ij} represents the area of each grid cell in the NIO region. The second term on the right-hand side denotes the mean nTA value for the NIO region between 1993 and 2020.

2.3.7. Empirical Orthogonal Function (EOF) Analysis

We employ Empirical Orthogonal Function (EOF) analysis to understand the effect of climate modes like El-Niño Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) on interannual variability of the surface TA in the AS and the BoB region separately. We remove a 5-year running climatology from surface TA and divide it by a 5-year running standard deviation to normalize the TA anomalies (Equation 16) (Murata et al., 2024; Valsala et al., 2020).

$$TA'(t) = \frac{TA(t) - \overline{TA}_{5yr}(t)}{\sigma_{5yr}(t)} \quad (16)$$

This process of calculating the normalized TA anomalies removes the long-term nonlinear trends and the seasonal cycle using a 5-year running mean anomaly, which substantially reduces the total variance in the TA data. While standardization is done to avoid regional dominance due to variance in magnitude differences. Thus, 5-year running anomalies act as a low-pass filter, allowing for retaining low-frequency ENSO/IOD-related signals.

To understand the effect of ENSO, we use the Nino3.4 index to correlate with the first and second principal components of TA (PC-1 and PC-2). The difference between the SST and the climatology over the central Pacific region (5°N – 5°S , 170° – 120°W) is known as the Nino3.4 index. Similarly, the Dipole Moment Index (DMI) is used to determine the effect of IOD. The DMI represents the differences in SST anomalies between the western (50° – 70°E and 10°S – 10°N) and the southeastern part of the Indian Ocean (90° – 110°E and 10° – 0°S).

3. Results and Discussions

3.1. Exploratory Data Analysis and XGB Performance

As mentioned in Section 2.2, we extracted the predictors from the reanalysis product at the collocated ship-based surface TA observations. Hence, it is essential to understand the differences between observations and reanalysis data, as it may be the source of input uncertainty (as mentioned in Section 2.3.3). In this study, we made a comparison of SST and SSS extracted from reanalysis products with collocated ship-based SST and SSS observations. Approximately 50% (565 out of 1,144 data points) of the total ship-based observations have temperature and salinity profiles collected using the CTD package. The rest of the profiles are available at standard

depths, not at 1 m regular depth intervals. Therefore, we did not proceed with the calculation of MLD using ship-based observations in this study. To ensure the consistency of data available from two different sources (reanalysis product and ship-based observations), we check the relationship (correlation (r) as a proxy for the relationship) between the dependent and each of the independent state variables for both data sets.

Figure 5a depicts a comparison between the ship-based SST observations and the reanalysis SST. We observe a good agreement between the SST from reanalysis (COP_SST) and ship-based observations (Obs_SST). Both of these SSTs seem to be well correlated ($r = 0.84$ ($p < 0.01$)) with relatively small differences (RMSE = 1.06°C). However, from Figure 5a, we see that low values of Obs_SST (minimum = 20.96°C) are not well captured by COP_SST (minimum = 23.59°C). This occurs in the upwelling zone at the western coast of the AS, a region where decreasing SST is associated with increased SSS (Figure S3 in Supporting Information S1). However, such points are very few and located very near the coast. Since reanalysis data is an outcome of the model (Dréville et al., 2022), it may show such discrepancies at locations near the coast. From Figure 5b we observe that a good correlation exists between the SSS from reanalysis (COP_SSS) and ship-based observations (Obs_SSS) ($r = 0.73$ ($p < 0.01$)) but has a high RMSE (1.58). It is evident that low Obs_SSS are not well captured in COP_SSS. These low SSS values are primarily from the head bay and near-coast regions of the BoB. The freshwater from the river decreases the SSS in the head bay region. The long-term salinity data from rivers are unavailable, so the models show high SSS in the BoB. Although the COP_SSS is an assimilated product, due to the paucity of coastal SSS observations, the model-simulated reanalysis SSS values are not as low as the observations. The difference between the magnitude of COP_SSS and Obs_SSS data, specifically along the coast, is expected to result in high uncertainties in the coastal region.

Figure 5c shows that the surface TA has a weak inverse relationship with Obs_SST ($r = -0.31$ ($p < 0.01$)). COP_SST also shows a similar relation with surface TA ($r = -0.24$ ($p < 0.01$)). As the SST rises, the CO_2 solubility decreases; therefore, the ocean's ability to absorb more atmospheric CO_2 decreases. Since the impact of SST on TA is not direct but shows a proxy to the biological processes, a weaker relationship may be due to low nutrient availability at the surface. Figure 5d shows substantial variability in surface TA with changes in SSS. With an increase in Obs_SSS, the surface TA also increases ($r = 0.82$ ($p < 0.01$)). The COP_SSS ($r = 0.71$ ($p < 0.01$)) also reproduces the relationship between Obs_SSS and surface TA well.

Figure 5e depicts the variation of MLD with surface TA. Since MLD is not available from the ship-based observation, we use the MLD from the reanalysis product (COP_MLD). The MLD seems to have a positive relation with TA ($r = 0.32$ ($p < 0.01$)). Depending on the mixing strengths, the deeper waters with high TA reach the near-surface, which ultimately increases surface TA. This initial analysis indicates that the dependent and independent variables used in this study are of good quality. Therefore, we now proceed with training the XGB algorithm.

In order to avoid overfitting, we employ the k-fold ($k = 10$) cross-validation technique while training the XGB model with the training data set. The independent input variables are then normalized before the algorithm is trained. The hyperparameters of the XGB algorithm (Table 1) are tuned, and finally, the algorithm is trained. Figure 6 shows the performance of the XGB algorithm with the train and test sets separately. Almost similar correlations ($r = 0.91$ and 0.90 with train and test set, respectively) and RMSEs (RMSE = 49.36 and $50.32 \mu\text{mol kg}^{-1}$ with train and test set, respectively) show that the model is well trained and can be used for creating the best estimate using an ensemble approach (Section 2.3.2). The best estimated surface TA data will now be referred to as INCOIS_TA.

We carried out SHapley Additive exPlanations (SHAP) (Lundberg & Lee, 2017) analysis to quantify the relative contribution of each predictor variable on surface TA. Figure 7 represents a SHAP summary plot with SHAP values ($\mu\text{mol kg}^{-1}$) on the X-axis. The SHAP values quantify the effect of each independent predictor (SST, SSS, and MLD) on the surface TA in the NIO. The SHAP summary plot in Figure 7 illustrates the impact of each predictor on the model's surface TA. The color gradient indicates a linear increase of each of the predictor values from low (blue) to high (red). The SSS is the primary driver affecting the surface TA, which is well established in literature (Metzl et al., 2024; Millero et al., 1998). The high MLD values show positive SHAP values in Figure 7, indicating the increase in TA associated with the deepening of MLD. However, a small magnitude of negative SHAP values is associated with lower MLD but moderately high SSS. This shows that at some locations, the presence of moderately high SSS and the low mixing strength results in low surface TA values. The sub-surface

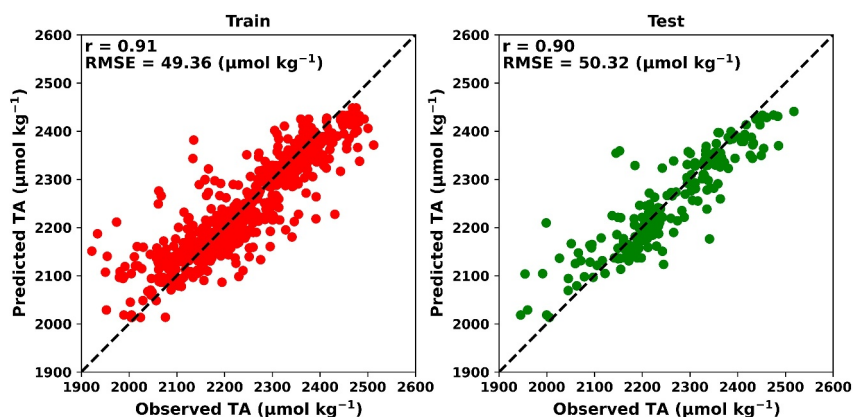


Figure 6. The performance of the trained Extreme gradient boosting algorithm against the train and test sets.

waters with high TA do not reach the near surface when the mixing strength is low (shallow MLD). From Figure 7, SST seems to have the lowest impact on the surface TA.

3.2. Comparison With Other Global Long-Term Surface TA Products

INCOIS_TA is a regional high-resolution data product. We assess its ability to capture the surface TA variability in the NIO with other recently developed global data products. This assessment is expected to help us understand if the performance of INCOIS_TA data product is consistent with other available data products. For the purpose of assessment, we have considered CMEMS-LSCE-FFNN (CMEMS_TA, (Chau et al., 2024)) data product which has spatial resolution of 0.25° and OceanSODAETHZ (SODA_TA, (Gregor & Gruber, 2021)) having spatial resolution of 1° . SODA_TA uses an ML approach to predict surface TA, whereas CMEMS_TA uses a regression approach to predict surface TA, but it has a higher spatial resolution. Further, we have employed another two algorithms to predict TA (ESPER_LIR and ESPER_NN (Carter et al., 2021)) in this assessment.

Figure 8 shows the probability distribution function (PDF) of surface TA from ship-based observations, two ML-based global products (CMEMS_TA and SODA_TA), ESPERs (ESPER_LIR and ESPER_NN), and INCOIS_TA for the NIO from 1993 to 2020. The PDF of ship-based observations shows that the surface TA in the NIO has values as low as $1979 \mu\text{mol kg}^{-1}$. These low surface TA are primarily found in the head bay region of BoB (due to the presence of low SSS). Compared to the observations, the lower surface TA values of CMEMS_TA, SODA_TA, ESPERs, and INCOIS_TA are restricted to $\approx 2000\text{--}2100 \mu\text{mol kg}^{-1}$. INCOIS_TA and ESPER_LIR has similar minimum surface TA of 2012 and $2030 \mu\text{mol kg}^{-1}$, whereas CMEMS_TA has $2084 \mu\text{mol kg}^{-1}$, ESPER_NN has $2074 \mu\text{mol kg}^{-1}$, and SODA_TA has $2121 \mu\text{mol kg}^{-1}$.

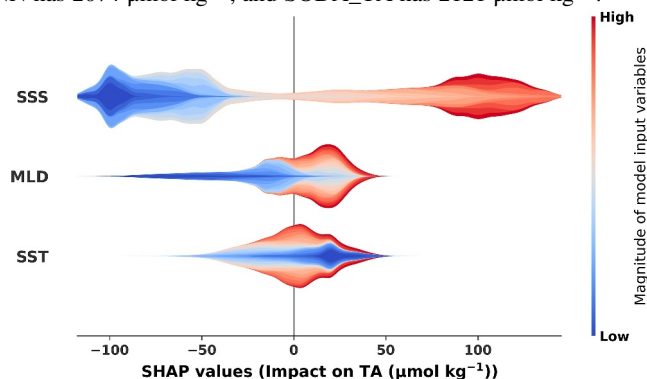


Figure 7. SHapley Additive exPlanations (SHAP) summary analysis showing the effect of each predictor on the surface total alkalinity (TA). The SHAP values quantify the impact of each predictor (sea surface temperature, sea surface salinity, and mixed layer depth). The greater the magnitude of the SHAP values, the stronger the impact. Positive SHAP values indicate an increase in surface TA, while negative values indicate a decrease. The color bar from blue to red represents a linear increase in the magnitude of each predictor.

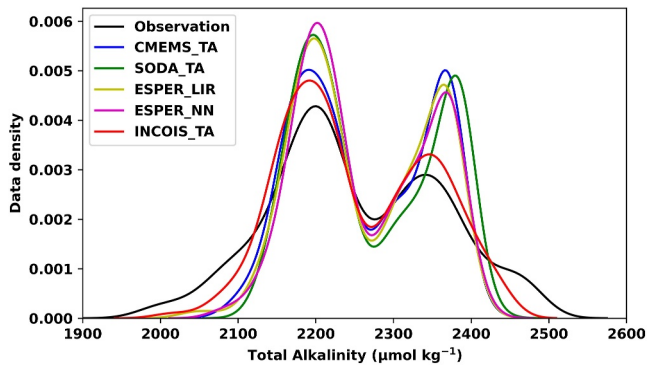


Figure 8. Data density distribution of predicted total alkalinity (TA) from various data products at the locations where TA observations are available.

Similarly, the ship-based observations have a maximum value of surface TA as $2492 \mu\text{mol kg}^{-1}$ (Figure 8). For the reconstructed data products (INCOIS_TA, CMEMS_TA, SODA_TA, and ESPERs), INCOIS_TA shows the maximum surface TA of $2439 \mu\text{mol kg}^{-1}$, which is found to be the closest. The high surface TA is usually found in the northwestern AS with high SSS resulting from strong upwelling. The PDFs of observations and different products show two peaks and one trough. The first peak of ship-based observations shows the surface TA to have a maximum frequency near $2200 \mu\text{mol kg}^{-1}$. In comparison, INCOIS_TA's first peak shows the maximum frequency to shift to a lower surface TA, but it comes closer to the observation's peak when compared with the other two global data products. The first peak of the reconstructed surface TA global data products (CMEMS_TA and SODA_TA) and ESPERs is found to be very high, which means that these products overestimate most of the surface TA values near $2200 \mu\text{mol kg}^{-1}$. INCOIS_TA also overestimates surface TA, but the

magnitude of overestimation is less than ESPERs, CMEMS_TA, and SODA_TA. For INCOIS_TA, the difference in the magnitude of reanalysis and observed SSS, as shown in Section 2.3.1.1, could be the reason for overestimation.

The trough and the second peak of ship-based observations and INCOIS_TA are much closer. To quantify which reconstructed surface TA data product's PDF covers the ranges of observation PDF, we calculate Perkins' skill score (PSS (Perkins, 1983)) as shown in Table 2. INCOIS_TA shows a better statistical performance (lower RMSE and higher correlation) than the other two surface TA products (Table 2). The use of the XGB algorithm (as compared to the other algorithms such as ESPERs, LIARv2, and SVR) for developing INCOIS_TA, availability of ship-based observations from the Indian EEZ region, and high spatial resolution (capable of capturing the variability of surface TA induced by the mesoscale ocean processes which are missing in the low-resolution global models) of INCOIS_TA are the possible reasons behind the improvements in the prediction.

3.3. Uncertainty Quantification

We quantify the uncertainty associated with the prediction of surface TA at each grid cell for the period 1993–2020. A detailed description of each step involved in calculating the total uncertainty is provided in Section 2.3.3. The MeU ($5 \mu\text{mol kg}^{-1}$), model uncertainty, input uncertainty, and representation uncertainty are summed in quadrature at each grid cell to provide the spatiotemporal total uncertainty. Figure 9 displays the total uncertainty and its three spatio-temporally varying components for the total period (1993–2020).

Figure 9a shows higher model uncertainty in the AS than the BoB. The western, central, and southwestern region of the AS shows high model uncertainties ($\approx 20\text{--}25 \mu\text{mol kg}^{-1}$). This could be due to the availability of a limited number of observations in this region, as observed from Figure 1. The western coast of India also shows high model uncertainties, especially near the Gujarat and Kerala coasts. The uncertainty in the northeastern part (near the Gujarat coast) ranges from 21.3 to $27.3 \mu\text{mol kg}^{-1}$, and for the southeastern part (near the Kerala coast), it varies from 22 to $27 \mu\text{mol kg}^{-1}$. The coastal BoB region does not show high model uncertainty values as opposed to the coastal AS. This is probably due to the availability of more coastal observations in the BoB. The average model uncertainty in the BoB region is seen to be $19.4 \mu\text{mol kg}^{-1}$.

The input uncertainty (Figure 9b) is seen to be lower in the open oceans ($\approx 10.4\text{--}14.4 \mu\text{mol kg}^{-1}$) but increases towards the coastal regions ($\approx 15\text{--}26 \mu\text{mol kg}^{-1}$). The highest input uncertainties are observed in the head-bay region of the BoB. Since the input uncertainty is driven by the errors in predictors taken from the reanalysis data, the open oceans, where observations and satellite data are available, should decrease the errors in predictors. The regions near the coast do have data, but may not be available for

Table 2

Statistical Comparison of the Surface Total Alkalinity Between INCOIS_TA (CMEMS_TA, SODA_TA, ESPER_LIR, ESPER_NN) and Ship-Based Observations

	RMSE ($\mu\text{mol kg}^{-1}$)	r ($p < 0.01$)	PSS
INCOIS_TA	48.5	0.89	0.78
CMEMS_TA	69.7	0.77	0.70
SODA_TA	71.0	0.78	0.65
ESPER_LIR	69.6	0.77	0.71
ESPER_NN	70.1	0.77	0.72

Note. The statistics shown in this table include Root Mean Squared Error (RMSE), Correlation (r), and Perkins' Skill Score (PSS). The p -value ($p < 0.01$) indicates 99% significance level.

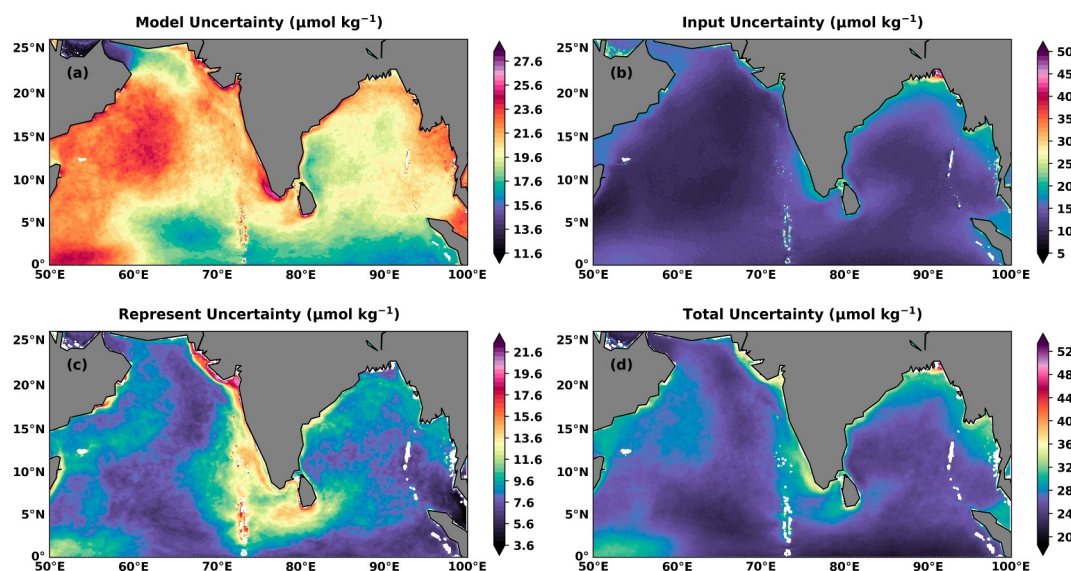


Figure 9. The figure shows (a) the model uncertainty, (b) the input uncertainty, and (c) the representative uncertainty. These three, together with measurement uncertainty (which are constant values added to each grid), form (d) the Total uncertainty. Each of these uncertainties is calculated per grid cell for the period 1993–2020 using Equation 10.

public use as they fall in the EEZ region. Hence, errors tend to increase towards the coastal regions.

The representative uncertainty is low in the AS ($<9 \mu\text{mol kg}^{-1}$) except for the eastern coastal part of AS (Figure 9c). The northeastern part (near the Gujarat coast) of AS shows higher representative uncertainty ($>16 \mu\text{mol kg}^{-1}$). Therefore, more observations may be required in the coastal part of eastern AS, especially in the northeastern part (near the Gujarat coast). The representative error in the BoB is almost the same throughout the domain ($\approx 10 \mu\text{mol/kg}$), except for the regions extremely near the west coast of the BoB. The higher representative uncertainty clearly indicates that we require high spatio-temporal resolution for capturing the characteristics of these highly variable regions.

Figure 9d represents the total uncertainty, which seems to be high in the coastal part of eastern AS and the coastal parts of northern BoB. It is evident that observational efforts to collect surface TA and associated essential ocean variables need to be enhanced in these regions. While we acknowledge that it is far simpler to advocate for the need for additional observations than it is to conduct cruises (especially when it comes to the logistics and resources), we nevertheless expect that the places with high uncertainty identified in this study will help to focus on the planning of future observational programs.

The differing methodologies used by various studies ((Chau et al., 2022, 2024; Gregor & Gruber, 2021)) to calculate total uncertainty make quantitative comparisons between the total uncertainty of INCOIS_TA and other studies challenging. For example, Chau et al. (2022, 2024) and Gregor and Gruber (2021) do not use an ensemble approach and hence do not calculate model uncertainty as done in this study. Moreover, Chau et al. (2024) provides uncertainty in predicted TA based on errors in the inputs used for prediction, while Gregor and Gruber (2021) does not have input uncertainty as a component of its total uncertainty in predicted TA. However, qualitative analysis reveals that other studies often show a steep gradient of uncertainty, with higher values in coastal regions compared to open-ocean areas. In contrast, INCOIS_TA substantially reduces this gradient, resulting in lower uncertainties in coastal regions. This improvement can likely be attributed to the integration of coastal observation data, the use of the XGB algorithm, and the high spatial resolution of INCOIS_TA.

3.4. Seasonal Variability of Surface TA

Figure 10 depicts a seasonal composite (seasonal average over 1993–2020) of the surface TA for the NIO from INCOIS_TA and two other global TA products (CMEMS_TA and SODA_TA). All the products capture the contrasting surface TA of the two ecosystems (AS and BoB) well. The pre-monsoon season (MAM) shows high

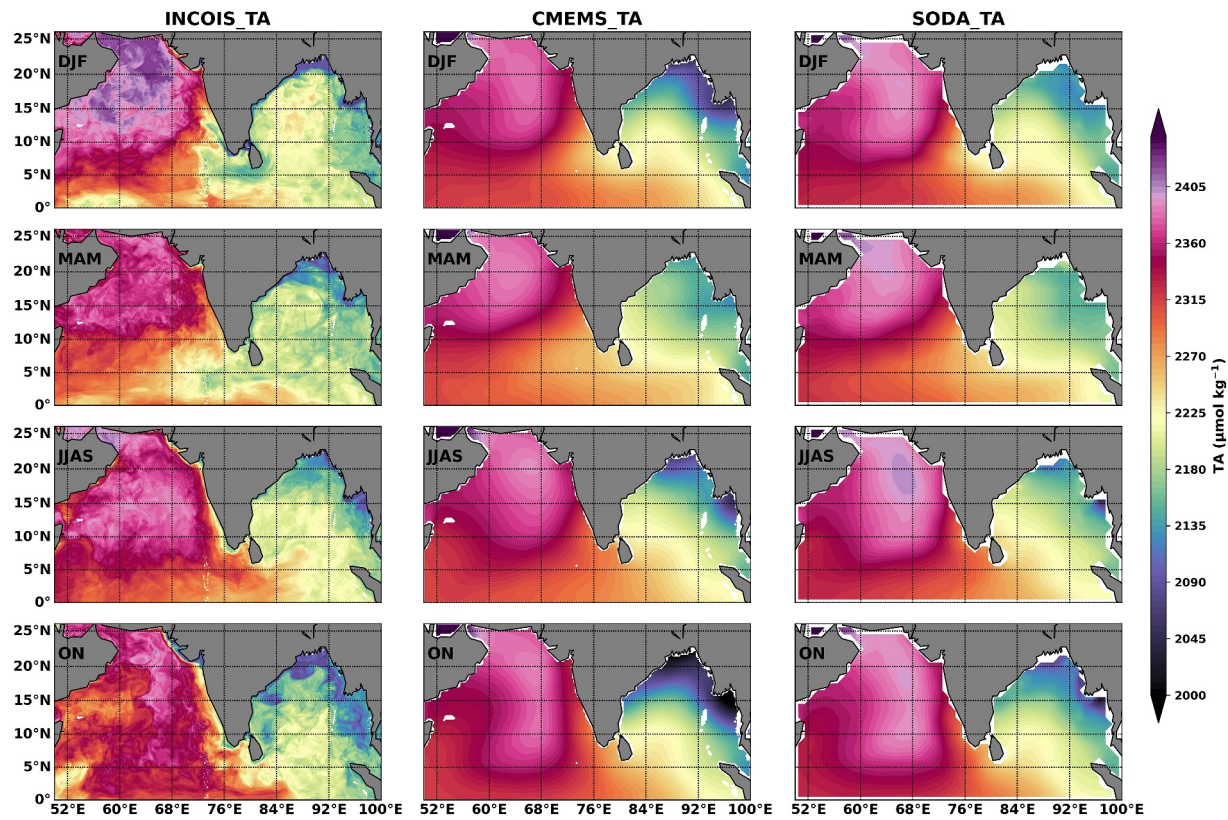


Figure 10. Seasonal composite of surface total alkalinity from INCOIS_TA is shown in the first column. The second and third column shows similar seasonal composites from CMEMS_TA and SODA_TA for the period 1993–2020. Each row of the figure represents a season.

surface TA in the northern AS, and the value of surface TA decreases as we move south. In the north of AS, INCOIS_TA (CMEMS_TA and SODA_TA) data product shows an average surface TA of 2352 (2381 and 2371) $\mu\text{mol kg}^{-1}$ during MAM, while for the southern AS, the average surface TA value is of 2298 (2307 and 2318) $\mu\text{mol kg}^{-1}$. INCOIS_TA shows lower surface TA values in comparison with the other two products. In the northern AS, high shortwave radiation during MAM increases SST and SSS (S. P. Kumar & Narvekar, 2005). The low winds during MAM are unable to break the stratification caused by high incoming radiation (S. P. Kumar & Narvekar, 2005), which inhibits further rise in SSS. As SSS is the primary driver of surface TA, environmental conditions such as high shortwave radiation and low winds influence surface TA in a manner similar to their influence on SSS. All the products (INCOIS_TA, CMEMS_TA, and SODA_TA) show low surface TA (2218, 2205, and 2203 $\mu\text{mol kg}^{-1}$) in the BoB, which is primarily due to the presence of lower SSS in the BoB than the AS.

Further, all the products show that during the southwest monsoon (June–September (JJAS)), surface TA in the AS region increases (Figure 10). The strong upwelling in the northwestern AS (due to high wind speeds and low evaporation (S. P. Kumar & Narvekar, 2005)) and the upwelling in the open oceans due to the Findlater Jet (Findlater, 1969) increases surface TA in the AS. In addition to this, the advection of TA-rich waters from the west to the east leads to the spread of high TA toward the east (Morrison et al., 1998; Sarma, 2003). In contrast, the monsoon precipitation and river discharge reduce SSS, increase the BoB stratification, and shallow MLD, resulting in a decrease in the surface TA. In the post-monsoon season (October–November (ON)), the southward-moving East India Coastal Current distributes the low-saline freshwater from north to south, which has a strong influence on the carbonate dynamics of the BoB (Joshi et al., 2021; Joshi & Warrior, 2022; Sarma et al., 2018). However, the values of surface TA decrease more in the northern BoB. During ON, in the AS, the strength of mixing reduces due to the presence of low winds (S. P. Kumar & Narvekar, 2005), which decreases the surface TA values. Figure 10 shows that only INCOIS_TA captures the decrease of surface TA values due to reduced

mixing. The use of MLD as a predictor makes it possible to consider the effects of mixing in the construction of surface TA.

All the models show maximum surface TA in the AS during the northeast monsoon season (December–February (DJF)) as shown in Figure 10. The influence of dry, low, northeasterly winds, low radiations, and high evaporation increases the SSS in the northern AS. These favorable conditions further initiate winter convection (S. P. Kumar & Narvekar, 2005), but convective processes dominate the surface, resulting in high TA. The BoB has a strong stratification due to the formation of a thick barrier layer in the northeast monsoon season (Joshi et al., 2021), which increases the retention time of freshwater at the sea surface. The stratification in the BoB has a strong control on the carbonate chemistry (increasing the surface TA) of the BoB region (Joshi et al., 2021; Joshi & Warrior, 2022). Strong variability in the TA and other inorganic carbon components with reference to coastal circulation associated with the redistribution of freshwater discharge was reported in the BoB (Sarma et al., 2018). Hence, the seasonal analysis indicates that INCOIS_TA can capture the seasonal variability similar to that of CMEMS_TA and SODA_TA in both the ecosystems (AS and BoB) of the NIO region.

3.5. Trend Analysis

3.5.1. Trends in the TA Predictors

Before analyzing the trends of surface TA, it is crucial to examine the trends of its predictors. Figures 11a–11c show the spatial trends of SST, SSS, and MLD between 1993 and 2020 (blue cross marks indicate regions having 90% significant trends). Warming of surface waters in the entire NIO was noticed, but the rate of warming is variable in different regions. For instance, a higher rate of warming was observed in the central, northernmost, and Somalia coasts in the AS and eastern BoB. Sridevi et al. (2023) reported spatial and seasonal variations in the warming trends associated with trends in anthropogenic aerosol optical depth (AOD) in the NIO. Sridevi et al. (2023) reported that the warming trends were not significant in the north of 12°N during winter, spring, and fall, associated with an increase in anthropogenic AOD. In contrast, the removal of anthropogenic AOD due to precipitation warms the entire NIO significantly during summer (Sridevi et al., 2023).

A significant increase in SSS (Figure 11b) is observed over almost the entire AS (except in the central AS) and the southern BoB. Sridevi and Sarma (2021) reported that winter convective mixing has been increasing in recent decades, caused by a decrease in the lower atmospheric temperature due to an increase in AOD that hinders solar radiation. Therefore, we observe SST ($\approx 0.017^{\circ}\text{C yr}^{-1}$) warming associated with increasing SSS (≈ 0.005 PSU yr^{-1}) in the northern AS. To the south of 5°N, the rate of SST increase is slightly higher ($0.02^{\circ}\text{C yr}^{-1}$), but the rate of SSS increase is significantly higher (0.01 PSU yr^{-1}) than that found in the northern AS regions. This higher rate of increase of both SSS and SST in the region south of 5°N could be attributed to an insignificant increase in anthropogenic AOD (Sridevi et al., 2023).

In the case of BoB, higher warming was observed in the eastern Bay, where anthropogenic AOD trends are weaker, associated with a decrease in salinity. Although the decreasing trends in salinity were not significant at 90% confidence levels (using the Wald test with t-distribution) between 1993 and 2020, the trends were significant between 2001 and 2020. The decrease in salinity in the BoB was attributed to an increase in river discharge due to the melting of Himalayan glaciers (Goes et al., 2020; Sridevi et al., 2023; Sridevi & Sarma, 2021). The strong stratification caused by increased freshwater discharge and weaker aerosol cover in the atmosphere led to a higher rate of warming in the eastern BoB (Figure 11a). The MLD trends in both basins are seen to be insignificant, except for the southeastern AS and a few regions in the southwestern BoB (Figure 11c).

3.5.2. Trends in TA

The TA displayed a significant increasing trend at the rate of $0.4\text{--}2.0 \mu\text{mol kg}^{-1} \text{yr}^{-1}$ at three sub-regions, namely south of 7°N, southeastern coast of the AS and the southwestern region of the BoB (Figure S4 in Supporting Information S1 shows the trends of domain-averaged surface TA in the AS and BoB regions, respectively), whereas insignificant trends are noticed in other parts of the study region (Figure 11d). The warming of the upper ocean is expected to reduce vertical nutrient supply and decrease biological calcification (Roxy et al., 2016), which may lead to an increase in TA in the NIO. Coccolithophores contribute to the primary productivity in the NIO region by $\approx 5\text{--}10\%$, and the contribution is higher in the AS than in the BoB (Moore et al., 2012).

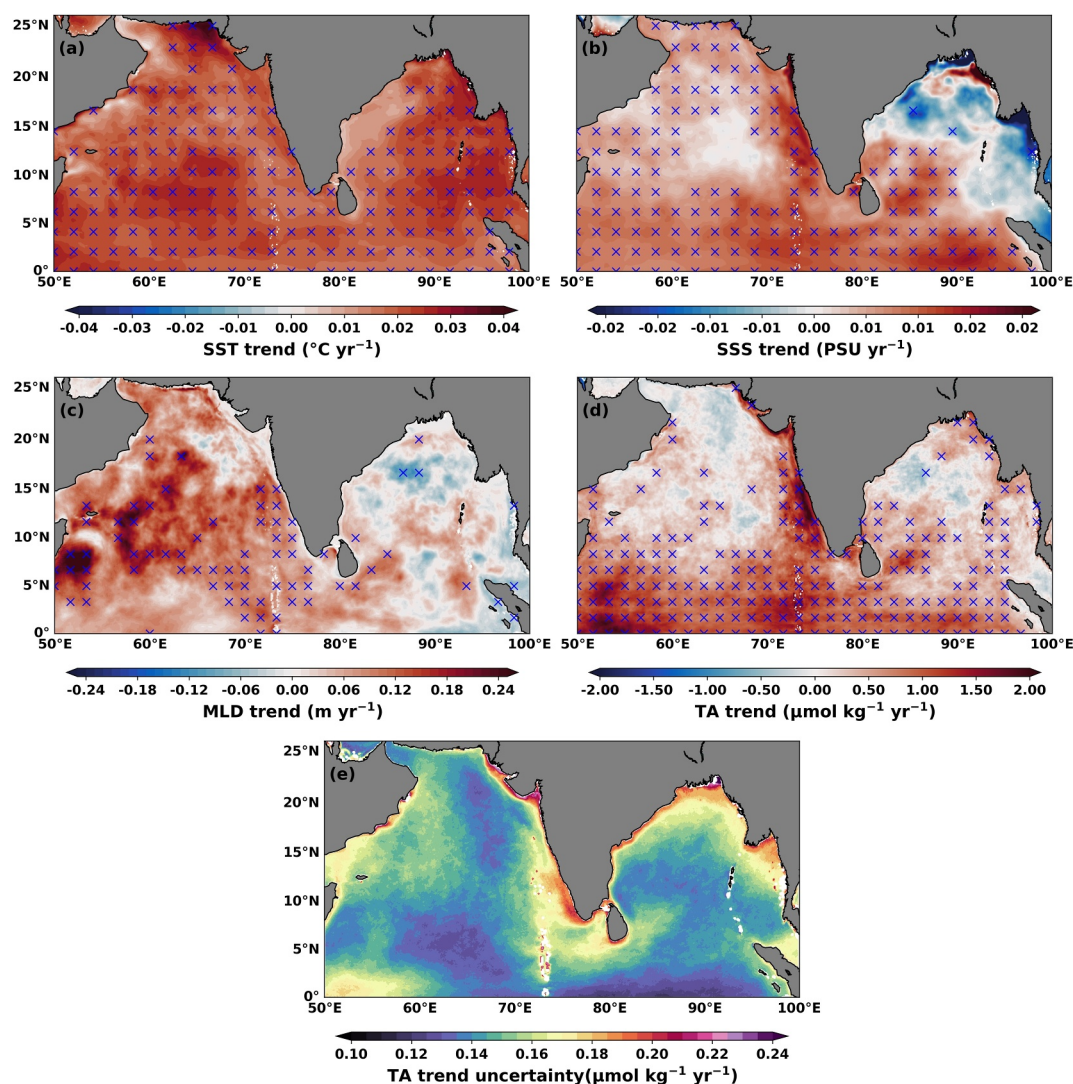


Figure 11. The figure represents the trend of (a) sea surface temperature ($^{\circ}\text{C yr}^{-1}$), (b) sea surface salinity (PSU yr^{-1}), (c) mixed layer depth (m yr^{-1}), and (d) total alkalinity (TA) ($\mu\text{mol kg}^{-1} \text{yr}^{-1}$) per grid cell over the whole study period and the corresponding p-values (blue cross marks) that are less than 0.1 (90% significance). Figure (e) shows the spatial uncertainty corresponding to the TA trends.

Figure 11e displays the uncertainty in the TA trend arising due to the uncertainties in the TA predictions (using the methodology described in Section 2.3.4). As expected, the spatial pattern of the uncertainty in the TA trend matches that of the ToU (Figure 9d) in TA predictions. The open ocean has low uncertainties in the TA trend ($0.12\text{--}0.15 \mu\text{mol kg}^{-1} \text{yr}^{-1}$). These uncertainties are observed to be highest ($>0.2 \mu\text{mol kg}^{-1} \text{yr}^{-1}$) near the eastern coast of the AS, western coast of the BoB, and the head bay regions of the BoB. To further analyze the probable reasons contributing to the TA trend, we quantify the effect of the trends in each of the input variables (SST, SSS, and MLD) on the TA trend (following the methodology described in Section 2.3.5).

Figure 12 displays the spatially averaged contribution of the trend in each of the input variables to the TA trend. We calculate the contribution for the regions having significant (Figure 12a) TA trend in each of the two ecosystems of the NIO (AS and BoB). The regions showing significant TA trend in the NIO regions are primarily driven by the trends in SSS in both basins (AS (70.3%) and BoB (53.4%)). The contributions of the trends SST (18.5%) and MLD (14.8%) are almost similar in the AS, but the trend of MLD (5.8%) has a negligible contribution to the TA trend of the BoB.

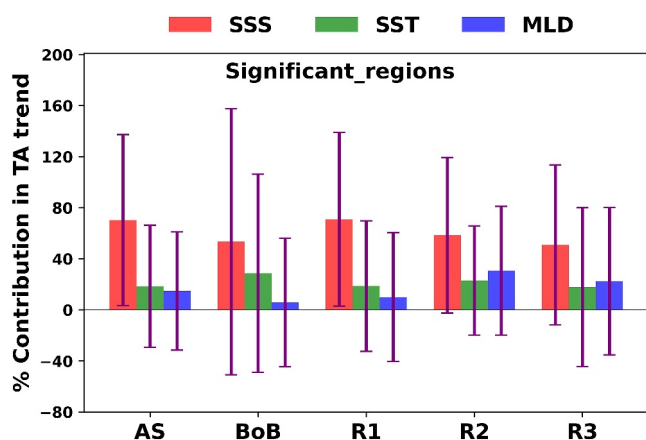


Figure 12. The figure represents the spatially averaged percentage contribution (with bars showing spatial standard deviation) of trends in each predictor (sea surface temperature, sea surface salinity, and mixed layer depth) to the total alkalinity (TA) trend in the regions having a significant (p -value < 0.1) TA trend. While referring to the two ecosystems, we refer to them as Arabian Sea (AS) and Bay of Bengal (BoB). However, R1 (region south of 7°N), R2 (the southeastern coast of the AS), and R3 (the southwestern region of BoB) refer to the three specific zones with significant TA trend.

The region south of 7°N (R1) shows a significant TA trend (Figure 11d) and spreads over both the AS and BoB regions. The contribution of input trends to the TA trend in this region is primarily driven by SSS (70.8%) (Figure 12, R1). In addition, Sridevi et al. (2023) observed a decline in net primary production (NPP) south of 7°N , whereas no significant NPP trends were found north of 7°N . This suggests that reduced carbonate uptake may contribute to the increase in TA in the south of 7°N . Generally, equatorial regions of the Pacific and Atlantic Oceans are sources for atmospheric CO_2 due to the occurrence of equatorial upwelling (Feely et al., 2002; Monteiro et al., 2022). In contrast, the lack of this feature in the Indian Ocean (Hood et al., 2024), along with rising TA levels, suggests that the equatorial Indian Ocean may currently absorb more atmospheric CO_2 , with this uptake potentially increasing in the future due to the ongoing TA trend. Although similar to the south of 7°N , increasing trends in SSS primarily control the increasing TA trends in the southeastern coast of the AS (R2, by 58.4%) and the southwestern region of the BoB (R3, by 50.9%) (Figure 12), the increasing trend in MLD contributes (30.6% in R2, and 22.5% in R3) more than SST (22.9% in R2 and 17.7% in R3) in these two regions (Figure 12).

Despite having significant SSS and SST trends, the TA trend is insignificant in most of the regions north of 7°N . The lack of significant trends (Figure 11d) of surface TA may be due to the deposition of atmospheric acidic pollutants (like nitrate and sulfate) that may neutralize the increase in TA in the northern

AS, where a significant increase in anthropogenic AOD is reported (Sridevi et al., 2023). Kumari et al. (2022) conducted aerosol dissolution experiments in the NIO region and found that they are acidic in nature, with an increase in acidity during winter and spring (November to May) compared to other seasons. While warming increased stratification and reduced vertical nutrient inputs in this region (Roxy et al., 2016), these effects were offset by the deposition of atmospheric nutrients (Sarma et al., 2022; Sridevi et al., 2023). Further, using numerical model outputs, Malsang et al. (2024) reported that over 70% of the nutrient loss from reduced vertical mixing in the NIO regions is compensated by atmospheric deposition. As a result, the impact of warming on net primary productivity in this region has been negligible over the past two decades, with little to no effect on calcification rates. This could be a potential reason for the lack of significant TA trends north of 7°N in the NIO region.

Sarma et al. (2021) noticed that the rate of ocean acidification is three times higher in the northern coastal BoB due to the deposition of atmospheric pollutants (like nitrate and sulfate). Kumari et al. (2021, 2022) found that the atmospheric deposition effect is more in the northern Bay but decreases toward the south due to the dilution of atmospheric pollutants. This may result in a weaker trend of the TA in the north and north-central of BoB (Figure 11d). Also, the Ganges discharge is relatively alkaline due to the flow of water through carbonaceous rocks in the Himalayan region. The pH (NBS scale) of Ganges waters of 8.130 ± 0.24 was reported at the Hooghly estuary (Mukhopadhyay et al., 2002), which is higher than that of the BoB. This may negate the effect of low nutrients at the surface due to increased stratification. Since TA is primarily influenced by SSS and in the northern BoB, the SSS trend is insignificant (Figure 11b), this likely explains why there is no significant TA trend in that region. Similar to the regions north of 7°N of the NIO, the South China Sea also receives a significant amount of anthropogenic pollutants from the atmosphere (Zhang & Reid, 2010), and similar changes may be possible, whereas excess TA may be titrated with deposition of acids, leading to insignificant trends in TA.

The increase in TA in the upper ocean allows absorption of more atmospheric CO_2 and enhances sequestration (Zhu et al., 2024). This is positive feedback from climate change that allows for the removal of more atmospheric CO_2 into the ocean. Sarma (2002) first reported an increase in TA at the depth of aragonite saturation horizon in the global ocean, except in the Atlantic, due to dissolution of calcium carbonate driven by the dissolution of anthropogenic CO_2 . However, trends in surface TA were not reported due to a lack of enough data to construct such trends. This study suggests that increasing trends in TA are noticed in the regions that are influenced by significant warming associated with a decrease in nutrient input to the surface and have insignificant deposition of atmospheric pollutants. The NIO (especially northern AS and BoB) receives the highest atmospheric pollutants in the world, and the rate of increase is also the highest (Sridevi et al., 2023; Yadav et al., 2021; Zhang &

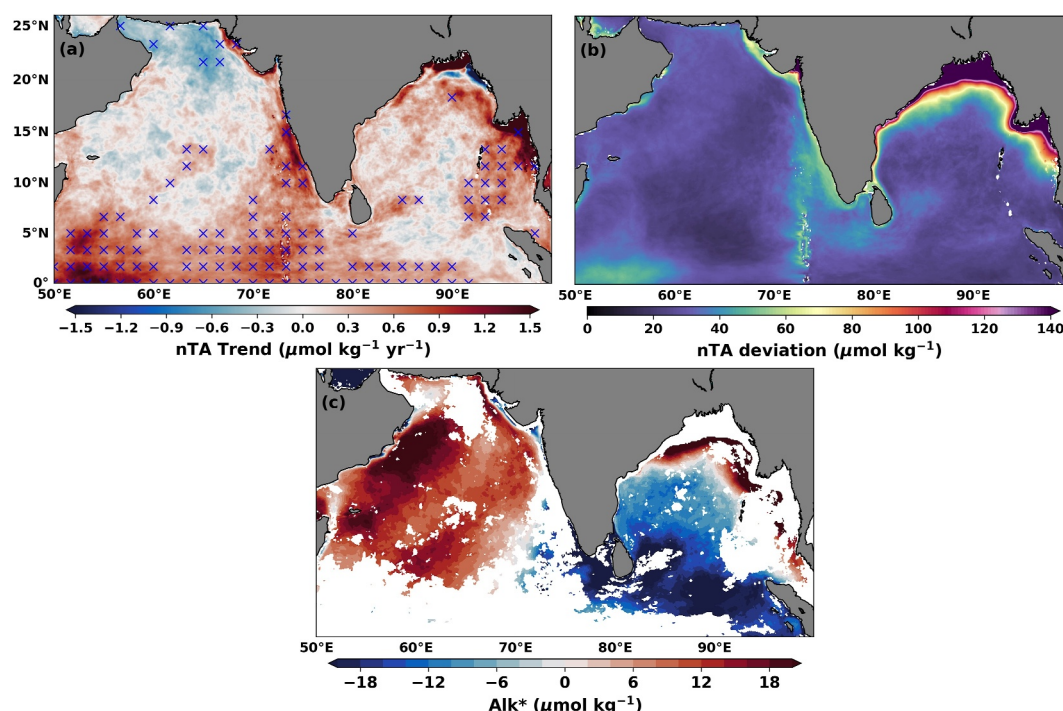


Figure 13. Spatial (a) mean trend of the Normalized TA (nTA) and the corresponding p-values (blue crosses) less than or equal to 0.1 (which represents 90% significance). Figure (b) shows the spatial standard deviation ($1-\sigma$) of nTA. Figure (c) shows annual mean values of Alk*. The white spaces in Figure (c) are areas removed in which total alkalinity is not normalized.

Reid, 2010). Further, tropical regions, excluding the NIO, are experiencing significant warming. Bindoff et al. (2022) projected a decline in primary productivity between the 1990s and 2100 in the tropical ocean due to climate change. These regions are also less affected by atmospheric pollutant deposition. In addition, nutrient availability is a major factor controlling primary production. As a result, an increase in TA may be expected due to reduced calcification rates. Therefore, this study suggests that TA may increase in tropical regions (excluding the NIO) due to a decline in primary production, potentially enhancing the ocean's buffering capacity.

Moreover, the TA trends observed in the Indian Ocean could have potential linkages to TA cycling in other global ocean basins, primarily driven by ocean circulation and biogeochemical processes. The Indian Ocean exchanges water with the Pacific Ocean mainly through the Indonesian Throughflow, and with the Atlantic Ocean via the Agulhas Current system (Gordon, 1985; Sprintall et al., 2014). These ocean gateways serve as major pathways by which water masses carrying distinct TA signatures, shaped by local processes in the Indian Ocean, could be transported to adjacent basins. This implies that changes in TA within the Indian Ocean may propagate beyond the basin boundaries and could influence the carbonate chemistry of connected oceans (Wiggert et al., 2013).

3.5.3. Normalized TA (nTA)

Figure 13a shows the trends in nTA. We observe that the regions that show a significant nTA trend (marked by blue crosses) are similar to the regions having a significant TA trend (Figure 11d). Thus, these regions may not have any effect of normalization because after normalization, the value nTA is similar to that of TA (Figure S5 in Supporting Information S1). So, the regions showing significant nTA trends could arise due to the trend in TA.

Figure 13b shows the standard deviation of nTA over time at each grid cell. Ideally, effective normalization should yield a uniform standard deviation across the NIO domain. In Figure 13b, we observe that while most regions exhibit similar variability, high values are found in the east coast of the AS and the head bay region of BoB. These regions are influenced by the zero salinity river alkalinity from the freshwater rivers (especially in the northern part of the BoB). It also suggests that TA_{river} of $840 \mu\text{mol kg}^{-1}$ still isn't able to completely remove the impact of zero salinity riverine TA. Further, TA_{river} exhibits distinct characteristics in the AS and BoB. Although

we have not split the data set, splitting the data may produce better normalization and lead to more accurate results by accounting for regional differences in riverine influence across the two basins. So, from Figures 13a and 13b, it is evident that normalization is ineffective in some regions, where TA and nTA remain similar and the riverine influence persists. Hence, before calculating Alk^* , we remove grid cells with nTA standard deviation exceeding $3\text{-}\sigma$ (assuming it reflects the excess zero salinity riverine TA).

The spatial annual mean pattern of Alk^* (Figure 13c) shows that its magnitude is close to zero in the NIO region. Similarly, Fry et al. (2015) showed the magnitude of Alk^* in the tropical regions to be close to zero. The mean nTA in the NIO region from our TA estimate is $2289\text{ }\mu\text{mol kg}^{-1}$. The nTA mean values reported by Fry et al. (2015) in the Indian Ocean were $2294\text{ }\mu\text{mol kg}^{-1}$, which is very close to the values estimated in this study. The spatial pattern of Alk^* reflects the ocean's carbon cycle (Fry et al., 2015). The positive values of Alk^* in the AS are caused by the strong inorganic pump, which brings the excess sub-surface TA to the surface. However, strong stratification in the BoB weakens the inorganic pump, reducing the TA at the surface (similar mechanisms affect the nutrient cycle in the NIO region (S. P. Kumar et al., 2002)). The positive Alk^* in the AS may indicate favorable calcification conditions, while vice versa in the BoB. However, positive Alk^* in the northern part BoB indicates that even after removing regions having $>3\text{-}\sigma$ nTA, it may not have removed the zero salinity riverine TA from the northern BoB.

3.6. Interannual Variability of Surface TA

In this section, we conduct the EOF analysis, as outlined in Section 2.3.6, to examine the influence of ENSO and IOD on the interannual variability of surface TA in the NIO. The analysis is performed separately for the two ecosystem zones of the NIO (AS and BoB). Figures 14a and 14b illustrate the first and second modes of the EOF's spatial pattern of the surface TA in the AS. The first mode explains about 7.05% while the second mode explains 5.58% of the total variance in space and time. The first mode of spatial EOF shows a basin-wide increase in surface TA (Figure 14a). Figure 14b shows a decrease in surface TA in the eastern AS and an increase in the west.

The PC-1 of the surface TA correlates inversely with the DMI and Nino3.4 indices (Figure 14c). The effect of ENSO ($r = -0.32$) is more dominant than IOD ($r = -0.22$). Although PC-1 exhibits apparent anomalies during the ENSO and/or IOD events (as explained in Section 2.3.7), its overall correlation during 1993–2020 remains weak. This reflects the limitations of linear correlation when applied to sparse, event-driven, and nonlinear climatic signals. With only a limited number of ENSO/IOD events during these 28 years (8 El Niño events, 10 La Niña events, and seven positive and one negative IOD events), statistical dilution due to the majority of the neutral years may reduce the correlation strength. Hence, in this study, we report the variability of PC-1 of TA during key ENSO and/or IOD events.

Sarma (2006) reported that the positive IOD that occurred in 1997–1998 may not have any effect on surface TA in the NIO. However, this analysis was based on the SSS observations available only in the southern Indian Ocean (Sarma, 2006). But the present study shows that surface TA reduces significantly during 1997–1998 primarily due to the strong El-Niño than the pIOD (Figure 14c).

The El-Niño reduces mixing and shallows the MLD in the AS (Singh et al., 2019), thus may inhibit the high sub-surface salinity and TA from reaching the surface, reducing the surface TA in the AS. During 2015 (weak pIOD but strong El-Niño), PC-1 shows a similar decrease in surface TA in the AS. During the La Niña periods, we do not observe noteworthy patterns in PC-1 of surface TA. PC-2 of TA shows relatively higher inverse correlation with ENSO ($r = -0.27$) than IOD ($r = -0.02$).

Figures 15a and 15b show spatial patterns of the first and second modes of EOF for surface TA in the BoB. The first mode explains 8.02% variance of surface TA, while the second mode explains 4.37%. The western BoB shows a decrease in surface TA, and the rest of the region (especially the south BoB) shows an increase (Figure 15a). However, the second mode of surface TA shows an increase in surface TA except for southwestern BoB (Figure 15b).

Figure 15c shows the correlation between PC-1 of surface TA with other climate indices (DMI and Nino3.4). We find that the PC-1 relates inversely (although moderately) with both DMI ($r = -0.32$) and Nino3.4 ($r = -0.37$) indices. Hence, the IOD and the ENSO have an almost equal impact on the interannual changes of surface TA in the BoB. As stated earlier, the weak correlation over the complete period can often underestimate individual ENSO or IOD events. From Figure 15c, it is observed that the surface TA decreased during the 1997–1998 pIOD and El-Niño period, while it increased in the 1998–1999 negative IOD (nIOD) and La Niña period. During 2002–

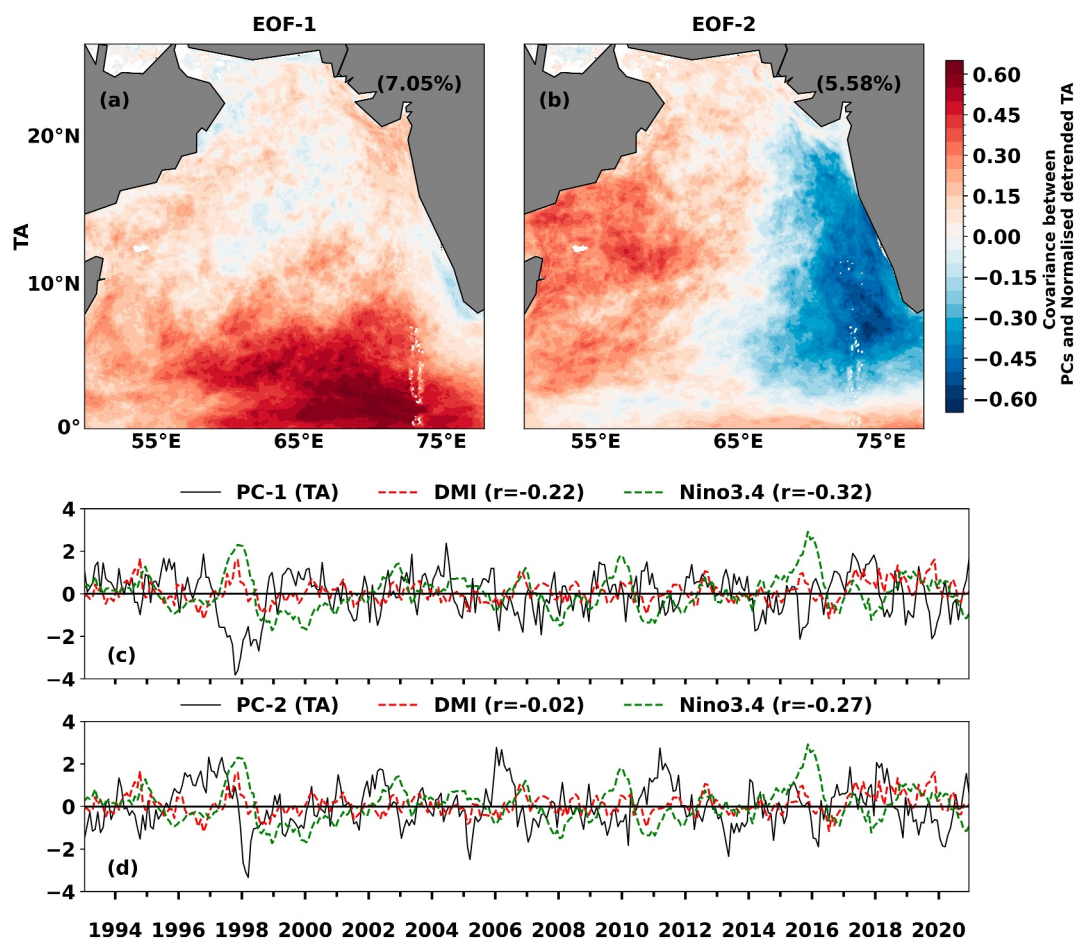


Figure 14. Figures (a) and (b) show the Empirical Orthogonal Function spatial patterns for total alkalinity (TA) in the Arabian Sea. The correlations of PC-1 and PC-2 of TA (c, d) with DMI and Niño3.4 indices indicate the effect of El Niño–Southern Oscillation and Indian Ocean Dipole.

2003, El-Niño, a slight decrease in surface TA is observed. Other strong El-Niño and pIOD periods like 2015–2016 and 2018–2019 also experienced a decrease in surface TA in the BoB region. The 2010–2011 had a strong La Niña, raising the surface TA of the BoB. The PC-2 of surface TA shows no effect of IOD or ENSO (Figure 15d). Both these climate indices affect the wind speeds, which further affect mixing in the BoB region. Thus, stratification or mixing could play a dominant role in controlling the variability of surface TA in the BoB.

The relatively low variance and overall weak correlation with the climatic events over the whole period may indicate that these modes only modestly reflect the interannual variability of TA. However, the impact of individual events is relatively well captured. We do not overinterpret these results but instead acknowledge these results as just initial observations. Thus, it warrants deeper analysis in the future, with better spatial recognition methods to explore the drivers of TA interannual variability in the NIO region.

As this study observes a moderate impact of climatic modes on TA in the NIO region, a similar influence on oceanic and atmospheric circulation patterns is expected, affecting precipitation, river runoff, and mixing processes across multiple basins (Du et al., 2009; Saji et al., 1999). These teleconnections can indirectly alter TA cycling in other regions by modifying the physical and chemical environment. Although the transfer of TA signals through these pathways occurs over timescales of months to decades, the regional impact is often most pronounced near oceanic gateways such as the Indonesian archipelago or the Agulhas region.

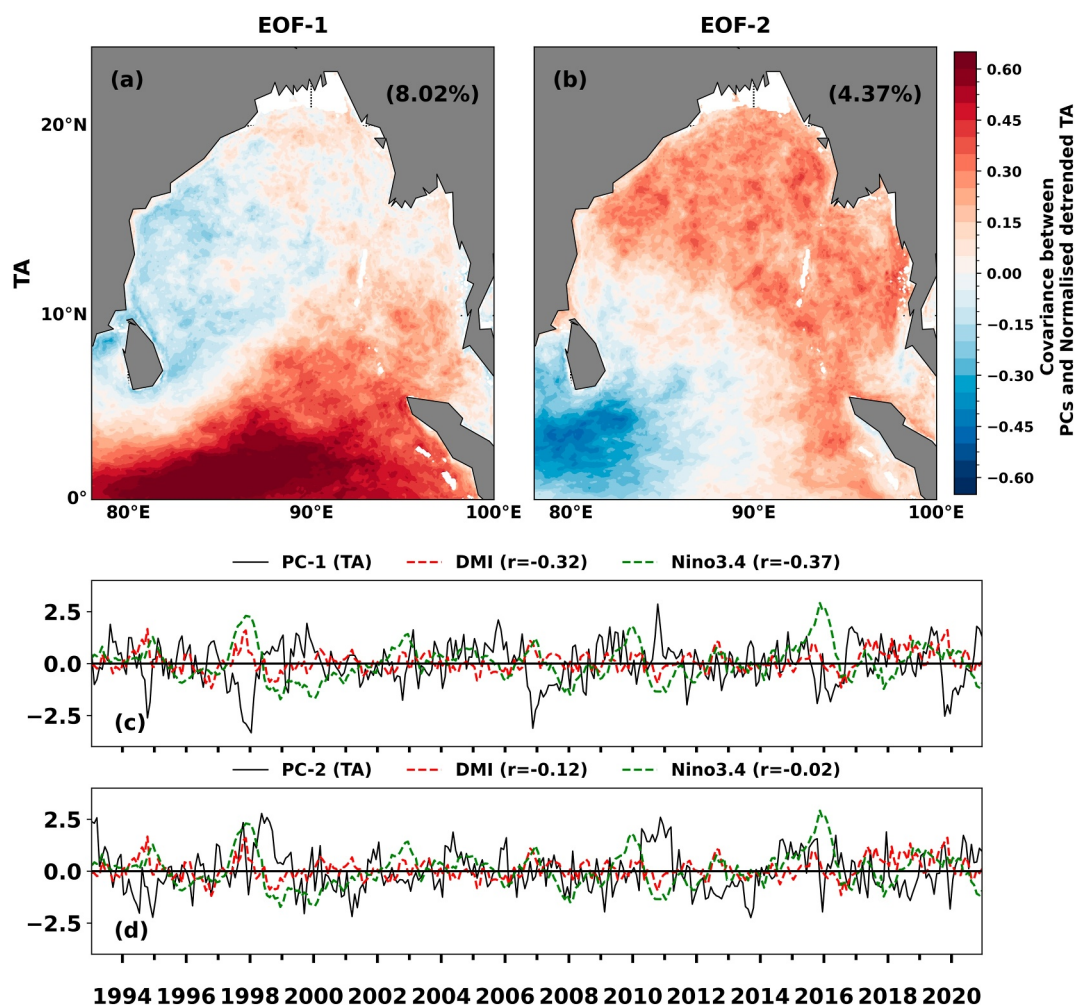


Figure 15. Figures (a) and (b) show the Empirical Orthogonal Function spatial patterns for total alkalinity (TA) in the Bay of Bengal. The correlations of PC-1 and PC-2 of TA (c, d) with DMI and Niño3.4 indices indicate the effect of El Niño–Southern Oscillation and Indian Ocean Dipole.

4. Summary and Conclusions

This study aims to reconstruct long-term (1993–2020) high-resolution ($1/12^\circ$) surface TA data product for the NIO to study its long-term changes in a changing environment. We assume that surface TA variability primarily depends on SST, SSS, and MLD changes. The ship-based surface TA observations were collected from different sources for the period 1978–2019, and collocated SST, SSS, and MLD were obtained from a reanalysis product to generate 140 XGB-based models using a tuned XGB algorithm.

An ensemble is created using the predictions of each of these 140 models to find out the best estimate of surface TA for the NIO, along with an estimation of spatio-temporal uncertainty. INCOIS_TA performs better than the two other ML-based global surface TA products. The availability of a significant number of coastal and open ocean observations makes it possible for the high-resolution INCOIS_TA to capture the region-specific TA variability resulting from the changes in ocean processes at the local and regional scales. Additionally, the use of the boosting ML algorithm and an ensemble approach further makes it possible to enhance the performance (as compared to ESPER_LIR and ESPER_NN) of INCOIS_TA in the NIO.

We identified four sources of uncertainty. The total uncertainty (i.e., the quadrature sum of the uncertainties from the identified four sources) is far less in the open oceans ($< 25 \mu\text{mol kg}^{-1}$) than in the coastal and northern BoB regions. The southeastern tip and the northeastern coast ($> 40 \mu\text{mol kg}^{-1}$) are found to have higher uncertainties.

Therefore, it is advisable to make observations with an increased sampling frequency in these high-uncertain regions to improve the performance of INCOIS_TA further.

The analysis of the seasonal variability of surface TA indicates that the AS's surface TA is $\approx 150\text{--}350\text{ }\mu\text{mol kg}^{-1}$ higher than the BoB's surface TA. SSS highly influences TA, and the SSS of AS is also higher than that of the BoB. The higher SSS in the AS is often attributed to upwelling and convective mixing. Thus, we may attribute the higher TA in the AS than in BoB to these processes. Similarly, freshwater flux and the associated stratification primarily influence the low SSS (and hence low TA) in the BoB. The lowest surface TA is seen in the northern BoB ($\approx 2050\text{--}2130\text{ }\mu\text{mol kg}^{-1}$).

The spatial trend analysis of the TA shows increased trends in the region south of 7°N , the southeast AS coast, and the southwestern region of the BoB. The quantification of the effect of the trends in input variables driving the significant TA trends indicates that the increasing SSS trend majorly contributes to the significant increasing TA trend in the AS (70.3%) and BoB (53.4%), respectively. After SSS, the trends of SST and MLD almost equally impact the significant TA trend in the AS (SST (18.5%) and MLD (14.8%)). However, the trend in MLD has minimal impact (5.8%) on the significant TA trend in the BoB region. The lack of a significant TA trend in the northern regions of NIO is associated with a significant trend in anthropogenic AOD, suggesting that atmospheric deposition of acidic pollutants may neutralize the TA trend in the northern region of NIO. The spatial pattern of the annual mean of excess alkalinity at the surface (Alk^*) suggests that the AS could be more favorable for calcification than the BoB.

The EOF analysis of the surface TA of the AS indicates ENSO (PC-1 correlation with Niño3.4 ($r = -0.32$)), more than IOD (PC-1 correlation with DMI ($r = -0.22$)), weakly impacts the TA in this region. The surface TA in the AS decreases during the strong El-Niño and strong pIOD period 1997–1998. The minimum surface TA in the AS is found during the 1997–1998 strong El-Niño and strong pIOD event. The maximum rise in surface TA is seen during the 2011–2012 La Niña event. The reduced mixing due to the El-Niño inhibits vertical transport of ocean tracers like salinity. This leads to a decrease in SSS in the AS. Given the strong correlation between salinity and TA, the mechanism reducing SSS may also reduce the surface TA.

The TA in the BoB is equally affected by both ENSO (PC-1 correlation with Niño3.4 ($r = -0.37$)) and IOD (PC-1 correlation with DMI ($r = -0.32$)). Thus, both climate indices have an almost equal impact on surface TA in the BoB region. Since these climate indices strongly impact mixing, stratification may be the major driver controlling surface TA in the BoB region. The stronger stratification during El-Niño and pIOD years may restrict the high sub-surface TA from reaching the surface, thus reducing the surface TA and vice versa during the La Niña and nIOD years. The low variance and overall weak correlation with the climatic events over the whole period may indicate that these modes only modestly reflect the interannual variability of TA. However, distinct PC-1 anomalies during individual climatic events indicate that their impacts are still appreciable. We avoid over-interpretation of these results and acknowledge that these results are just initial observations. Thus, a further analysis in the future is required to identify the potential drivers of the TA interannual variability in the NIO region.

About 30% of the world's population resides along the coastal regions of the Indian Ocean (Doyle, 2018; Ghoshal et al., 2025; A. Kumar et al., 2024; Wafar et al., 2011). Anthropogenic activities significantly impact ocean health in this region. Understanding the variability in TA is, therefore, crucial for assessing its role in buffering the rise in ocean acidification. Our study highlights the probable impact of atmospheric pollutant deposition on TA in the northern part of the NIO. This study provides valuable insights into how climate change feedback mechanisms help maintain stable surface water TA in the NIO. The observed increasing trends in surface TA suggest an enhancement in the buffering capacity of the NIO region. Consequently, TA dynamics may influence the long-term carbon storage capacity of the NIO, potentially affecting the global TA budget.

To the best of our knowledge, INCOIS_TA is the only regional high-resolution long-term (gap-free) surface TA product available for the NIO. The INCOIS_TA product could play a critical role in validating and quantifying uncertainties in global numerical and ML-based biogeochemical models for the NIO ecosystem. Furthermore, it could be used to bias-correct Coupled Model Intercomparison Project (CMIP) global models for the two ecosystems of the Indian Ocean. These bias-corrected CMIP models could then be applied to project future surface TA trends in the NIO under various scenarios. Such projections would improve our understanding of the NIO

region's contribution to the future global alkalinity budget, particularly in light of the high discharge of low-salinity and high-alkalinity river water into the Indian Ocean.

The reduction of error between the reanalysis and observed SSS and SST will further improve the performance of INCOIS_TA. Also, depending on the continuous increase in sampling frequency, INCOIS_TA will be updated in the future to improve upon the current predictions. Our future work will also focus on developing regional high-resolution carbonate variables with depth-varying information and probabilistic future predictions.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

GLODAP data (Lauvset et al., 2022) are freely available from <https://www.glodap.info/>. WOCE data from CLIVAR and Carbon Hydrographic Data Office (CCHDO, 2023) is available from <https://cchdo.ucsd.edu/>. OceanSODA data (Gregor & Gruber, 2021) can be obtained from <https://www.ncei.noaa.gov/data/oceans/ncei/ocads/data/0220059/>. CMEMS-LSCE-FFNN data (Chau et al., 2024) is downloaded from <https://doi.org/10.14768/a2f0891b-763a-49e9-af1b-78ed78b16982>. We downloaded the GLORY (Drévilion et al., 2022) data (SST, SSS, and MLD) from https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_PHY_001_030/description. The NO_3^{-1} (Perruche, 2018) is downloaded from https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_BGC_001_029/download?dataset=cmems_mod_glo_bgc_my_0.25_P1M-m. We downloaded Dipole Moment Index (DMI) data (Saji et al., 1999) from <https://www.jamstec.go.jp/virtualearth/general/en/index.html>. The NINO3.4 is taken from <http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices>. WOA18 (Garcia et al., 2019) is downloaded from <https://www.ncei.noaa.gov/data/oceans/woa/WOA18/DATA/>. EN4.2.2 data (Gouretski & Cheng, 2020) is available at <https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-2.html>. As per the Indian data-sharing policy, data collected within the Indian Exclusive Economic Zone (EEZ) cannot be made publicly available or posted on any website, even with password protection. Due to this restriction imposed by the Government of India, we are unable to submit EEZ cruise data to any open-access data portal. However, the data used in this study will be made available upon request by contacting the corresponding author at kcincoishyb@gmail.com or kunal.c@incois.gov.in. The surface TA data (along with total uncertainty) (Joshi, 2025) developed in this study is available at <https://zenodo.org/records/15645422>. This manuscript uses the colormaps from the Py-ART toolkit (Helmus & Collis, 2016). The Py-ART library can be downloaded from <https://arm-doe.github.io/pyart/>.

Acknowledgments

We are grateful to the associate editor and anonymous reviewers for their careful reading, constructive comments, and helpful suggestions, which have helped us to significantly improve the analysis and presentation of this work. INCOIS_TA data product has been developed as a part of the “Development of Climate Change Advisory Services” project of the Indian National Centre for Ocean Information Services, Hyderabad, India, under the “Deep Ocean Mission” program of the Ministry of Earth Sciences (MoES), Govt. of India. The authors would like to thank the contributions of various scientists involved with the GLODAPv2 project. Director, MoES-INCOIS and Director, CSIR-NIO are gratefully acknowledged for their encouragement to this work. RR and CJ acknowledge the Director, National Remote Sensing Centre (NRSC) Hyderabad, Director, Earth Observation & Disaster Management (EDPO), ISRO, CGM (RC's) and DGM, RRSC East, NRSC for their constant encouragement. This is INCOIS contribution number 584.

References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2623–2631).
- Barrett, R. C., Carter, B. R., Fassbender, A. J., Tilbrook, B., Woosley, R. J., Azetsu-Scott, K., et al. (2025). Biological responses to ocean acidification are changing the global ocean carbon cycle. *Global Biogeochemical Cycles*, 39(3), e2024GB008358. <https://doi.org/10.1029/2024gb008358>
- Bates, N. R., Pequignat, A. C., & Sabine, C. L. (2006). Ocean carbon cycling in the Indian Ocean: 1. Spatiotemporal variability of inorganic carbon and air-sea CO_2 gas exchange. *Global Biogeochemical Cycles*, 20(3). <https://doi.org/10.1029/2005gb002491>
- Bindoff, N. L., Cheung, W. W., Kairo, J. G., Aristegui, J., Guinder, V. A., Hallberg, R., et al. (2022). Changing ocean, marine ecosystems, and dependent communities.
- Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., et al. (2018). An alternative to static climatologies: Robust estimation of open ocean CO_2 variables and nutrient concentrations from T, S, and O_2 data using Bayesian neural networks. *Frontiers in Marine Science*, 5, 328. <https://doi.org/10.3389/fmars.2018.00328>
- Brewer, P. G., Bradshaw, A., & Williams, R. (1986). Measurements of total carbon dioxide and alkalinity in the North Atlantic Ocean in 1981. In *The changing carbon cycle: A global analysis* (pp. 348–370). Springer.
- Brewer, P. G., Wong, G. T., Bacon, M. P., & Spencer, D. W. (1975). An oceanic calcium problem? *Earth and Planetary Science Letters*, 26(1), 81–87. [https://doi.org/10.1016/0012-821x\(75\)90179-x](https://doi.org/10.1016/0012-821x(75)90179-x)
- Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., et al. (2019). A global monthly climatology of total alkalinity: A neural network approach. *Earth System Science Data*, 11(3), 1109–1127. <https://doi.org/10.5194/essd-11-1109-2019>
- Canadell, J. G., Monteiro, P. M., Costa, M. H., Cotrim da Cunha, L., Cox, P. M., Eliseev, A. V., et al. (2023). Intergovernmental Panel on Climate Change (IPCC). Global carbon and other biogeochemical cycles and feedbacks. In *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the Intergovernmental panel on climate change* (pp. 673–816). Cambridge University Press.
- Carter, B. R., Bittig, H. C., Fassbender, A. J., Sharp, J. D., Takeshita, Y., Xu, Y.-Y., et al. (2021). New and updated global empirical seawater property estimation routines. *Limnology and Oceanography: Methods*, 19(12), 785–809. <https://doi.org/10.1002/lom3.10461>

- Carter, B. R., Feely, R., Mecking, S., Cross, J., Macdonald, A., Siedlecki, S., et al. (2017). Two decades of Pacific anthropogenic carbon storage and Ocean acidification along Global Ocean Ship-based Hydrographic investigations Program sections P16 and P02. *Global Biogeochemical Cycles*, 31(2), 306–327. <https://doi.org/10.1002/2016gb005485>
- Carter, B. R., Feely, R., Williams, N., Dickson, A., Fong, M., & Takeshita, Y. (2018). Updated methods for global locally interpolated estimation of alkalinity, pH, and nitrate. *Limnology and Oceanography: Methods*, 16(2), 119–131. <https://doi.org/10.1002/lom3.10232>
- CCHDO. (2023). CCHDO hydrographic data archive [Dataset]. *UC San Diego Library Digital Collections*. <https://doi.org/10.6075/J0CCHAM8>
- Chau, T. T., Gehlen, M., & Chevallier, F. (2022). A seamless ensemble-based reconstruction of surface ocean $p\text{CO}_2$ and air–sea CO_2 fluxes over the global coastal and open oceans. *Biogeosciences*, 19(4), 1087–1109. <https://doi.org/10.5194/bg-19-1087-2022>
- Chau, T.-T., Gehlen, M., Metzl, N., & Chevallier, F. (2024). CMEMS-LSCE: A global, 0.25°, monthly reconstruction of the surface ocean carbonate system. *Earth System Science Data*, 16(1), 121–160. <https://doi.org/10.5194/essd-16-121-2024>
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794).
- Dickson, A. G. (1992). The development of the alkalinity concept in marine chemistry. *Marine Chemistry*, 40(1–2), 49–63. [https://doi.org/10.1016/0304-4203\(92\)90047-e](https://doi.org/10.1016/0304-4203(92)90047-e)
- Dickson, A. G., Sabine, C. L., & Christian, J. R. (2007). *Guide to best practices for ocean CO_2 measurements*. North Pacific Marine Science Organization.
- Dickson, A. G., & Whitfield, M. (1981). An ion-association model for estimating acidity constants (at 25° C and 1 atm total pressure) in electrolyte mixtures related to seawater (ionic strength <1 mol $\text{kg}^{-1}\text{H}_2\text{O}$). *Marine Chemistry*, 10(4), 315–333. [https://doi.org/10.1016/0304-4203\(81\)90012-8](https://doi.org/10.1016/0304-4203(81)90012-8)
- Doney, S. C., Busch, D. S., Cooley, S. R., & Kroeker, K. J. (2020). The impacts of ocean acidification on marine ecosystems and reliant human communities. *Annual Review of Environment and Resources*, 45(1), 83–112. <https://doi.org/10.1146/annurev-environ-012320-083019>
- Doney, S. C., Fabry, V. J., Feely, R. A., & Kleypas, J. A. (2009). Ocean acidification: The other CO_2 problem. *Annual Review of Marine Science*, 1(1), 169–192. <https://doi.org/10.1146/annurev.marine.010908.163834>
- Doyle, T. (2018). Blue economy and the Indian Ocean rim. *Taylor & Francis*, 14(1), 1–6. <https://doi.org/10.1080/19480881.2018.1421450>
- Drévillon, M., Lellouche, J.-M., Régnier, C., Garric, G., Bricaud, C., Hernandez, O., & Bourdallé-Badie, R. (2022). Global ocean reanalysis products global_reanalysis_phy_001_030.
- Du, Y., Xie, S.-P., Huang, G., & Hu, K. (2009). Role of air–sea interaction in the long persistence of El Niño–induced North Indian Ocean warming. *Journal of Climate*, 22(8), 2023–2038. <https://doi.org/10.1175/2008jcli2590.1>
- Feely, R. A., Boutin, J., Cosca, C. E., Dandonneau, Y., Etcheto, J., Inoue, H. Y., et al. (2002). Seasonal and interannual variability of CO_2 in the equatorial Pacific. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(13–14), 2443–2469. [https://doi.org/10.1016/s0967-0645\(02\)00044-9](https://doi.org/10.1016/s0967-0645(02)00044-9)
- Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., & Millero, F. J. (2004). Impact of anthropogenic CO_2 on the CaCO_3 system in the oceans. *Science*, 305(5682), 362–366. <https://doi.org/10.1126/science.1097329>
- Findlater, J. (1969). A major low-level air current near the Indian Ocean during the northern summer. *Quarterly Journal of the Royal Meteorological Society*, 95(404), 362–380. <https://doi.org/10.1002/qj.49709540409>
- Fine, R. A., Willey, D. A., & Millero, F. J. (2017). Global variability and changes in ocean total alkalinity from Aquarius satellite data. *Geophysical Research Letters*, 44(1), 261–267. <https://doi.org/10.1002/2016gl071712>
- Friis, K., Körtzinger, A., & Wallace, D. W. (2003). The salinity normalization of marine inorganic carbon chemistry data. *Geophysical Research Letters*, 30(2), 1085. <https://doi.org/10.1029/2002gl015898>
- Fry, C. H., Tyrrell, T., Hain, M. P., Bates, N. R., & Achterberg, E. P. (2015). Analysis of global surface ocean alkalinity to determine controlling processes. *Marine Chemistry*, 174, 46–57. <https://doi.org/10.1016/j.marchem.2015.05.003>
- Garcia, H., Boyer, T., Baranova, O., Locarnini, R., Mishonov, A., Grodsky, A. e., et al. (2019). In A. Mishonov (Ed.), *World ocean atlas 2018: Product documentation*.
- Ghoshal, P. K., Joshi, A., & Chakraborty, K. (2025). An improved long-term high-resolution surface $p\text{CO}_2$ data product for the Indian Ocean using machine learning. *Scientific Data*, 12(1), 577. <https://doi.org/10.1038/s41597-025-04914-z>
- Goes, J. I., Tian, H., Gomes, H. d. R., Anderson, O. R., Al-Hashmi, K., deRada, S., et al. (2020). Ecosystem state change in the Arabian Sea fuelled by the recent loss of snow over the Himalayan-Tibetan plateau region. *Scientific Reports*, 10(1), 1–8. <https://doi.org/10.1038/s41598-020-64360-2>
- Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data? *MIS Quarterly*, 36(3), 981–1001. <https://doi.org/10.2307/41703490>
- Gordon, A. L. (1985). Indian-Atlantic transfer of thermocline water at the Agulhas retroflection. *Science*, 227(4690), 1030–1033. <https://doi.org/10.1126/science.227.4690.1030>
- Gouretski, V., & Cheng, L. (2020). Correction for systematic errors in the global dataset of temperature profiles from mechanical bathythermographs. *Journal of Atmospheric and Oceanic Technology*, 37(5), 841–855. <https://doi.org/10.1175/jtech-d-19-0205.1>
- Goyet, C., Coatanoan, C., Eischeid, G., Amaoka, T., Okuda, K., Healy, R., & Tsunogai, S. (1999). Spatial variation of total CO_2 and total alkalinity in the northern Indian ocean: A novel approach for the quantification of anthropogenic CO_2 in seawater. *Journal of Marine Research*, 57(1), 135–163. <https://doi.org/10.1357/002224099765038599>
- Gregor, L., & Gruber, N. (2021). Oceansoda-ethz: A global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data*, 13(2), 777–808. <https://doi.org/10.5194/essd-13-777-2021>
- Helmus, J. J., & Collis, S. M. (2016). The python arm radar toolkit (py-art), a library for working with weather radar data in the python programming language. *Journal of Open Research Software*, 4(1), 25. <https://doi.org/10.5334/jors.119>
- Hood, R. R., Coles, V. J., Huggett, J. A., Landry, M. R., Levy, M., Moffett, J. W., & Rixen, T. (2024). Nutrient, phytoplankton, and zooplankton variability in the Indian Ocean. In *The Indian Ocean and its role in the global climate system* (pp. 293–327). Elsevier.
- Jiang, Z.-P., Tyrrell, T., Hydes, D. J., Dai, M., & Hartman, S. E. (2014). Variability of alkalinity and the alkalinity-salinity relationship in the tropical and subtropical surface ocean. *Global Biogeochemical Cycles*, 28(7), 729–742. <https://doi.org/10.1002/2013gb004678>
- Joshi, A. P. (2025). Long-term changes of surface total alkalinity in two contrasting basins of the North Indian Ocean [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.15645422>
- Joshi, A. P., Chowdhury, R. R., Warrior, H., & Kumar, V. (2021). Influence of the freshwater plume dynamics and the barrier layer thickness on the CO_2 source and sink characteristics of the Bay of Bengal. *Marine Chemistry*, 236, 104030. <https://doi.org/10.1016/j.marchem.2021.104030>
- Joshi, A. P., Ghoshal, P. K., Chakraborty, K., & Sarma, V. V. S. S. (2024). Sea-surface $p\text{CO}_2$ maps for the Bay of Bengal based on advanced machine learning algorithms. *Scientific Data*, 11(1), 384. <https://doi.org/10.1038/s41597-024-03236-w>

- Joshi, A. P., Kumar, V., & Warrior, H. V. (2022). Modeling the sea-surface $p\text{CO}_2$ of the central Bay of Bengal region using machine learning algorithms. *Ocean Modelling*, 178, 102094. <https://doi.org/10.1016/j.ocemod.2022.102094>
- Joshi, A. P., & Warrior, H. V. (2022). Comprehending the role of different mechanisms and drivers affecting the sea-surface $p\text{CO}_2$ and the air-sea CO_2 fluxes in the Bay of Bengal: A modeling study. *Marine Chemistry*, 243, 104120. <https://doi.org/10.1016/j.marchem.2022.104120>
- Karmakar, A., Parekh, A., Chowdary, J., & Gnanaseelan, C. (2018). Inter comparison of tropical Indian Ocean features in different ocean reanalysis products. *Climate Dynamics*, 51(1–2), 119–141. <https://doi.org/10.1007/s00382-017-3910-8>
- Kerr, D. E., Brown, P. J., Grey, A., & Kelleher, B. P. (2021). The influence of organic alkalinity on the carbonate system in coastal waters. *Marine Chemistry*, 237, 104050. <https://doi.org/10.1016/j.marchem.2021.104050>
- Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., et al. (2004). A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochemical Cycles*, 18(4). <https://doi.org/10.1029/2004gb002247>
- Key, R. M., Olsen, A., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., et al. (2015). Global ocean data analysis project, version 2 (GLODAP2). Ornl/Cdiac-162, Ndp-093.
- Krishna, K. V., & Shanmugam, P. (2023). Robust estimates of the total alkalinity from satellite oceanographic data in the global ocean. *IEEE Access*, 11, 42824–42838. <https://doi.org/10.1109/access.2023.3271516>
- Kumar, A., Susann, T., Fernandes, O. S., Biswas, H., Girach, I., Roxy, M. K., et al. (2024). Surface ocean-lower atmospheric processes in the Indian Ocean: Current understanding, knowledge gaps, and future directions. *Elementa: Science of the Anthropocene*, 12(1), 00041. <https://doi.org/10.1525/elementa.2023.00041>
- Kumar, S. P., Muraliedharan, P., Prasad, T., Gauns, M., Ramaiah, N., De Souza, S., et al. (2002). Why is the Bay of Bengal less productive during summer monsoon compared to the Arabian Sea? *Geophysical Research Letters*, 29(24), 88–1–88–4. <https://doi.org/10.1029/2002gl016013>
- Kumar, S. P., & Narvekar, J. (2005). Seasonal variability of the mixed layer in the central Arabian Sea and its implication on nutrients and primary productivity. *Deep Sea Research Part II: Topical Studies in Oceanography*, 52(14–15), 1848–1861. <https://doi.org/10.1016/j.dsr2.2005.06.002>
- Kumar, S. P., & Prasad, T. (1999). Formation and spreading of Arabian Sea high-salinity water mass. *Journal of Geophysical Research*, 104(C1), 1455–1464. <https://doi.org/10.1029/1998jc900022>
- Kumari, V., Sarma, V. V. S. S., & Kumar, M. D. (2022). Spatial variability in aerosol composition and its seawater acidification potential in coastal waters of the western coastal Bay of Bengal. *Journal of Earth System Science*, 131(4), 251. <https://doi.org/10.1007/s12040-022-01996-w>
- Kumari, V., Yadav, K., Sarma, V. V. S. S., & Dileep Kumar, M. (2021). Acidification of the coastal Bay of Bengal by aerosols deposition. *Journal of Earth System Science*, 130(4), 1–13. <https://doi.org/10.1007/s12040-021-01723-x>
- Lauvset, S. K., Lange, N., Tanhua, T., Bittig, H. C., Olsen, A., Kozyr, A., et al. (2022). GLODAPv2. 2022: The latest version of the global interior ocean biogeochemical data product. *Earth System Science Data*, 14(12), 5543–5572. <https://doi.org/10.5194/essd-14-5543-2022>
- Lee, K., Tong, L. T., Millero, F. J., Sabine, C. L., Dickson, A. G., Goyet, C., et al. (2006). Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. *Geophysical Research Letters*, 33(19), L19605. <https://doi.org/10.1029/2006gl027207>
- Lenton, A., Metzl, N., Takahashi, T., Kuchinke, M., Matear, R. J., Roy, T., et al. (2012). The observed evolution of oceanic $p\text{CO}_2$ and its drivers over the last two decades. *Global Biogeochemical Cycles*, 26(2). <https://doi.org/10.1029/2011gb004095>
- Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., et al. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, 2(12), 831–836. <https://doi.org/10.1038/ngeo689>
- Le Quéré, C., Takahashi, T., Buitenhuis, E. T., Rödenbeck, C., & Sutherland, S. C. (2010). Impact of climate change and variability on the global oceanic sink of CO_2 . *Global Biogeochemical Cycles*, 24(4). <https://doi.org/10.1029/2009gb003599>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- Malsang, M., Resplandy, L., Bopp, L., Zhao, Y., Ditkovsky, S., Yang, F., et al. (2024). Contemporary decline in northern Indian Ocean primary production weakly offset by rising atmospheric nitrogen deposition. *Frontiers in Marine Science*, 11, 1418634. <https://doi.org/10.3389/fmars.2024.1418634>
- Metzl, N., Fin, J., Lo Monaco, C., Mignon, C., Alliouane, S., Antoine, D., et al. (2024). A synthesis of ocean total alkalinity and dissolved inorganic carbon measurements from 1993 to 2022: The SNAPO-CO2-v1 dataset. *Earth System Science Data*, 16(1), 89–120. <https://doi.org/10.5194/essd-16-89-2024>
- Mignot, A., Claustre, H., Cossarini, G., D'Ortenzio, F., Gutknecht, E., Lamouroux, J., et al. (2023). Using machine learning and Biogeochemical-argo (BGC-argo) floats to assess biogeochemical models and optimize observing system design. *Biogeosciences*, 20(7), 1405–1422. <https://doi.org/10.5194/bg-20-1405-2023>
- Millero, F. J., Lee, K., & Roche, M. (1998). Distribution of alkalinity in the surface waters of the major oceans. *Marine Chemistry*, 60(1–2), 111–130. [https://doi.org/10.1016/s0304-4203\(97\)00084-4](https://doi.org/10.1016/s0304-4203(97)00084-4)
- Monteiro, T., Batista, M., Henley, S., Machado, E. d. C., Araujo, M., & Kerr, R. (2022). Contrasting Sea-Air CO_2 exchanges in the western tropical Atlantic Ocean. *Global Biogeochemical Cycles*, 36(8), e2022GB007385. <https://doi.org/10.1029/2022gb007385>
- Moore, T. S., Dowell, M. D., & Franz, B. A. (2012). Detection of coccolithophore blooms in ocean color satellite imagery: A generalized approach for use with multiple sensors. *Remote Sensing of Environment*, 117, 249–263. <https://doi.org/10.1016/j.rse.2011.10.001>
- Morrison, J. M., Codispoti, L., Gaurin, S., Jones, B., Manghnani, V., & Zheng, Z. (1998). Seasonal variation of hydrographic and nutrient fields during the US JGOFS Arabian Sea process study. *Deep Sea Research Part II: Topical Studies in Oceanography*, 45(10–11), 2053–2101. [https://doi.org/10.1016/s0967-0645\(98\)00063-0](https://doi.org/10.1016/s0967-0645(98)00063-0)
- Mukhopadhyay, S., Biswas, H., De, T., Sen, S., & Jana, T. (2002). Seasonal effects on the air–water carbon dioxide exchange in the Hooghly estuary, NE coast of Bay of Bengal, India. *Journal of Environmental Monitoring*, 4(4), 549–552. <https://doi.org/10.1039/b201614a>
- Murata, A., Kouketsu, S., Sasaoka, K., & Arulanathan, K. (2024). Modulation of surface seawater CO_2 system at 80°E: Impacts of the positive IOD in 2019. *Journal of Geophysical Research: Oceans*, 129(12), e2024JC021177. <https://doi.org/10.1029/2024jc021177>
- Olsen, A., Key, R. M., Van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., et al. (2016). The Global Ocean Data Analysis Project version 2 (GLODAPv2)—an internally consistent data product for the world ocean. *Earth System Science Data*, 8(2), 297–323. <https://doi.org/10.5194/essd-8-297-2016>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Perkins, K. (1983). On the use of composition scoring techniques, objective measures, and objective tests to evaluate ESL writing ability. *Tesol Quarterly*, 17(4), 651–671. <https://doi.org/10.2307/3586618>
- Perruche, C. (2018). Product user manual for the global ocean biogeochemistry hindcast global_reanalysis_bio_001_029. version 1.

- Prasad, T., & McClean, J. L. (2004). Mechanisms for anomalous warming in the western Indian Ocean during dipole mode events. *Journal of Geophysical Research*, 109(C2), C02019. <https://doi.org/10.1029/2003jc001872>
- Roxy, M. K., Modi, A., Murtugudde, R., Valsala, V., Panickal, S., Prasanna Kumar, S., et al. (2016). A reduction in marine primary productivity driven by rapid warming over the tropical Indian Ocean. *Geophysical Research Letters*, 43(2), 826–833. <https://doi.org/10.1002/2015gl066979>
- Saji, N., Goswami, B. N., Vinayachandran, P., & Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, 401(6751), 360–363. <https://doi.org/10.1038/43854>
- Sarma, V. V. S. S., Gupta, S., Babu, P., Acharya, T., Harikrishnachari, N., Vishnuvardhan, K., et al. (2009). Influence of river discharge on plankton metabolic rates in the tropical monsoon driven Godavari estuary, India. *Estuarine, Coastal and Shelf Science*, 85(4), 515–524. <https://doi.org/10.1016/j.ecss.2009.09.003>
- Sarma, V. V. S. S., Prasad, M., & Dalabehera, H. (2021). Influence of phytoplankton pigment composition and primary production on $p\text{CO}_2$ levels in the Indian Ocean. *Journal of Earth System Science*, 130(2), 1–16. <https://doi.org/10.1007/s12040-021-01598-y>
- Sarma, V. V. S. S. (2002). An evaluation of physical and biogeochemical processes regulating the oxygen minimum zone in the water column of the Bay of Bengal. *Global Biogeochemical Cycles*, 16(4), 46–146–10. <https://doi.org/10.1029/2002gb001920>
- Sarma, V. V. S. S. (2003). Monthly variability in surface $p\text{CO}_2$ and net air-sea CO_2 flux in the Arabian Sea. *Journal of Geophysical Research*, 108(C8), 3255. <https://doi.org/10.1029/2001jc001062>
- Sarma, V. V. S. S. (2006). The influence of Indian Ocean Dipole (IOD) on biogeochemistry of carbon in the Arabian Sea during 1997–1998. *Journal of Earth System Science*, 115(4), 433–450. <https://doi.org/10.1007/b02702872>
- Sarma, V. V. S. S., Kumar, M. D., & George, M. (1998). The central and eastern Arabian Sea as a perennial source of atmospheric carbon dioxide. *Tellus B: Chemical and Physical Meteorology*, 50(2), 179–184. <https://doi.org/10.3402/tellusb.v50i2.16095>
- Sarma, V. V. S. S., Kumari, V., Srinivas, T., Krishna, M., Ganapathi, P., & Murty, V. (2018). East India Coastal Current controls the dissolved inorganic carbon in the coastal Bay of Bengal. *Marine Chemistry*, 205, 37–47. <https://doi.org/10.1016/j.marchem.2018.07.010>
- Sarma, V. V. S. S., & Narvekar, P. V. (2001). A study on inorganic carbon components in the Andaman Sea during the post monsoon season. *Oceanologica Acta*, 24(2), 125–134. [https://doi.org/10.1016/s0399-1784\(00\)01133-6](https://doi.org/10.1016/s0399-1784(00)01133-6)
- Sarma, V. V. S. S., Prasad, V., Kumar, B., Rajeev, K., Devi, B., Reddy, N., & Kumar, M. D. (2010). Intra-annual variability in nutrients in the Godavari estuary, India. *Continental Shelf Research*, 30(19), 2005–2014. <https://doi.org/10.1016/j.csr.2010.10.001>
- Sarma, V. V. S. S., Sridevi, B., Kumar, A., Bikkina, S., Kumari, V., Bikkina, P., et al. (2022). Impact of atmospheric anthropogenic nitrogen on new production in the northern Indian ocean: Constrained based on satellite aerosol optical depth and particulate nitrogen levels. *Environmental Sciences: Processes & Impacts*, 24(10), 1895–1911. <https://doi.org/10.1039/d2em00234e>
- Sarmiento, J. L., Dunne, J., Gnanadesikan, A., Key, R., Matsumoto, K., & Slater, R. (2002). A new estimate of the CaCO_3 to organic carbon export ratio. *Global Biogeochemical Cycles*, 16(4), 54–1–54–12. <https://doi.org/10.1029/2002gb001919>
- Sarmiento, J. L., & Gruber, N. (2006). *Ocean biogeochemical dynamics*. Princeton university press.
- Sasse, T., McNeil, B., & Abramowitz, G. (2013). A novel method for diagnosing seasonal to inter-annual surface ocean carbon dynamics from bottle data using neural networks. *Biogeosciences*, 10(6), 4319–4340. <https://doi.org/10.5194/bg-10-4319-2013>
- Sauzède, R., Bittig, H. C., Claustre, H., Pasqueron de Fommervault, O., Gattuso, J.-P., Legendre, L., & Johnson, K. S. (2017). Estimates of water-column nutrient concentrations and carbonate system parameters in the global ocean: A novel approach based on neural networks. *Frontiers in Marine Science*, 4, 128. <https://doi.org/10.3389/fmars.2017.00128>
- Shaik, I., Yadav, S., Krishna, G., Mahesh, P., Nagamani, P. V., Begum, S., et al. (2024). Global surface ocean total alkalinity estimation: The machine learning approach. *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5. <https://doi.org/10.1109/lgrs.2024.3398048>
- Singh, S., Valsala, V., Prajeesh, A., & Balasubramanian, S. (2019). On the variability of Arabian Sea mixing and its energetics. *Journal of Geophysical Research: Oceans*, 124(11), 7817–7836. <https://doi.org/10.1029/2019jc015334>
- Sprintall, J., Gordon, A. L., Koch-Larrouy, A., Lee, T., Potemra, J. T., Pujiana, K., & Wijffels, S. E. (2014). The Indonesian Seas and their role in the coupled ocean–climate system. *Nature Geoscience*, 7(7), 487–492. <https://doi.org/10.1038/ngeo2188>
- Sridevi, B., Sabira, S., & Sarma, V. V. S. S. (2023). Impact of ocean warming on net primary production in the northern Indian ocean: Role of aerosols and freshening of surface ocean. *Environmental Science and Pollution Research*, 30(18), 53616–53634. <https://doi.org/10.1007/s11356-023-26001-9>
- Sridevi, B., & Sarma, V. (2024). Rapid shoaling of aragonite saturation horizon in the northern Indian Ocean. *Environmental Research Communications*, 6(5), 051006. <https://doi.org/10.1088/2515-7620/ad45c1>
- Sridevi, B., & Sarma, V. V. S. S. (2021). Role of river discharge and warming on ocean acidification and $p\text{CO}_2$ levels in the Bay of Bengal. *Tellus B: Chemical and Physical Meteorology*, 73(1), 1–20. <https://doi.org/10.1080/16000889.2021.1971924>
- Takahashi, T., Broecker, W. S., & Bainbridge, A. E. (1981). The alkalinity and total carbon dioxide concentration in the world oceans. Carbon cycle modelling. *Scientific Committee on Problems of the Environment*, 16(3078), 271–286.
- Takahashi, T., Sutherland, S. C., Chipman, D. W., Goddard, J. G., Ho, C., Newberger, T., et al. (2014). Climatological distributions of pH, $p\text{CO}_2$, total CO_2 , alkalinity, and CaCO_3 saturation in the global surface ocean, and temporal changes at selected locations. *Marine Chemistry*, 164, 95–125. <https://doi.org/10.1016/j.marchem.2014.06.004>
- Unesco. (1969). *Discharge of selected rivers of the world*. Author.
- Valsala, V., Sreeush, M., & Chakraborty, K. (2020). The IOD impacts on the Indian Ocean carbon cycle. *Journal of Geophysical Research: Oceans*, 125(11), e2020JC016485. <https://doi.org/10.1029/2020jc016485>
- Vinayachandran, P. N. M., Masumoto, Y., Roberts, M. J., Huggett, J. A., Halo, I., Chatterjee, A., et al. (2021). Reviews and syntheses: Physical and biogeochemical processes associated with upwelling in the Indian Ocean. *Biogeosciences*, 18(22), 5967–6029. <https://doi.org/10.5194/bg-18-5967-2021>
- Wafar, M., Venkataraman, K., Ingole, B., Ajmal Khan, S., & LokaBharathi, P. (2011). State of knowledge of coastal and marine biodiversity of Indian Ocean countries. *PLoS One*, 6(1), e14613. <https://doi.org/10.1371/journal.pone.0014613>
- Wiggert, J. D., Hood, R. R., Naqvi, S. W. A., Brink, K. H., & Smith, S. L. (2013). *Indian Ocean biogeochemical processes and ecological variability*. John Wiley & Sons.
- Wolf-Gladrow, D. A., Zeebe, R. E., Klaas, C., Körtzinger, A., & Dickson, A. G. (2007). Total alkalinity: The explicit conservative expression and its application to biogeochemical processes. *Marine Chemistry*, 106(1–2), 287–300. <https://doi.org/10.1016/j.marchem.2007.01.006>
- Wu, Z., Lu, W., Roobaert, A., Song, L., Yan, X.-H., & Cai, W.-J. (2024). A machine-learning reconstruction of sea surface $p\text{CO}_2$ in the North American Atlantic coastal ocean margin from 1993 to 2021. *Earth System Science Data Discussions*, 2024, 1–32.
- Yadav, K., Rao, V., Sridevi, B., & Sarma, V. V. S. S. (2021). Decadal variations in natural and anthropogenic aerosol optical depth over the Bay of Bengal: The influence of pollutants from Indo-Gangetic Plain. *Environmental Science and Pollution Research*, 28(39), 55202–55219. <https://doi.org/10.1007/s11356-021-14703-x>

- Zhang, J., & Reid, J. (2010). A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and Level 2 MISR aerosol products. *Atmospheric Chemistry and Physics*, 10(22), 10949–10963. <https://doi.org/10.5194/acp-10-10949-2010>
- Zhu, T., Zheng, L., Li, F., Liu, J., & Zhuang, W. (2024). Sustainable carbon sequestration via olivine based ocean alkalinity enhancement in the east and South China Sea: Adhering to environmental norms for nickel and chromium. *Science of the Total Environment*, 930, 172853. <https://doi.org/10.1016/j.scitotenv.2024.172853>