A Solely Satellite-Based Approach to Monitoring the Oceanic Carbonate System

William Feng

Abstract. In an attempt to develop a solely satellite-based approach to monitoring the oceanic carbonate system (OCS), a systematic and global-scale study was performed to evaluate a variety of machine learning models on their abilities to predict four OCS parameters, i.e., pH, total alkalinity (TA), dissolved inorganic carbon (DIC), and partial pressure of CO_2 (pCO_2) based solely on satellite-derived data. Random Forest (RF) was identified as the best performing model. When further validated against several publicly available *in situ* OCS datasets, RF demonstrated high predictive capability, achieving an R^2 of 0.95, 0.87, 0.68 and RMSE of 11.7, 16.8, and 0.025 for predicting DIC, TA, and pH respectively over the Global Ocean Data Analysis Project dataset, and achieving an R^2 of 0.70 and RMSE of 18.6 for predicting pCO_2 over the Lamont-Doherty Earth Observatory Global Surface pCO_2 dataset. This study has demonstrated the feasibility of monitoring the OCS based solely on satellite data and has identified RF as a useful machine learning model for estimating these OCS parameters at good accuracy, potentially decreasing our reliance on expensive and environmentally harmful *in situ* sampling.

1. Introduction

Ever since the start of the Industrial Revolution in around 1750, the average pH of ocean surface waters worldwide have been declining, in a process commonly referred to as "ocean acidification (OA)." OA is primarily caused by elevated levels of atmospheric carbon dioxide (CO_2), a direct result of human activities including fossil fuel burning, deforestation, and various industrial (e.g. petrochemical production, etc.), agricultural (e.g. livestock raising), and architectural (e.g. cement production) practices. As the level of atmospheric CO_2 rises, a strikingly high proportion (>30%) of it is taken up by oceans [1, 2]. Upon dissolution, CO_2 reacts with seawater to form carbonic acid, which in turn increases the acidity of the seawater.

Accompanying the declining oceanic pH in the OA process, carbonate ion concentrations ($[CO_3^{2-}]$) in the oceans have also been decreasing [1, 2, 3, 4], and as a direct impact, it is more

difficult for marine calcifying organisms (e.g. corals, mollusks) to build and maintain their calcium carbonate shells and skeletal structures [5, 6, 7]. Take corals as an example. Laboratory and field-based studies have indicated a direct, negative impact of OA on coral calcification [8]. It is estimated that OA alone could account for up to a 20.3% reduction in the skeletal density of reef-building corals in at least some regions of the oceans by the end of the 21st century [9]. Such declines in coral skeletal density cause increased susceptibility of coral reefs, widely recognized as the most biologically diverse ecosystems, to damages from pounding waves, storms, and other forms of erosion.

Currently, four key parameters that define the oceanic carbonate system (OCS), i.e., pH, total alkalinity (TA), dissolved inorganic carbon (DIC), and partial pressure of CO_2 (pCO_2), are commonly studied to assess OA, and can be further utilized to track changes in the ocean's carbonate chemistry and to evaluate the broader impacts of global climate change.

Today's methods to monitor OCS parameters rely heavily on *in situ* measurements, including the traditional manual approaches such as vessel-based water sampling or fixed station-based regular measurements, and have recently expanded to the sensor-based autonomous approaches such as moored buoys, profiling floats, underwater or surface gliders or other autonomous vehicles [10]. Although these *in situ* based approaches have provided crucial measurements for monitoring the OCS parameters, there are several noticeable disadvantages, including: (1) high costs, which include the typically expensive deployment, operation, maintenance, personnel, and/or fuel costs; (2) low spatiotemporal resolution (e.g. sparse geographic coverage and time series data), due to the various economical, technical, and operational constraints; and (3) unavoidable disturbance to the marine environment, which ironically include the emission of large amounts of CO_2, resulted typically from burning of fossil fuels in research or deployment vessels [11]. These limitations have caused large areas of the oceans to remain substantially unsampled, and many monitoring efforts to be substantially cost-prohibitive, and may further lead to systematic biases, such as the seasonal biases in data collection [12, 13].

In light of these limitations associated with the traditional *in situ* based monitoring methods, recently there have been efforts utilizing satellite-based remote sensing for the estimation of OCS parameter(s). For example, Sarma et al. demonstrated that pCO_2 in the North Pacific ocean can be estimated using satellite-derived sea surface temperatures (SSTs),

chlorophyll-a concentrations, and climatological salinity [14]. Lee et al. also reported the estimation of TA using satellite derived SST and sea surface salinity (SSS) data for different oceanographic regions [15]. Gregor et al. recently reported the development of OceanSODA (Satellite Oceanographic Datasets for Acidification)-ETHZ dataset aiming to utilize reanalysis data (a combination of satellite and *in situ* data) to model OA at a monthly resolution [16]. These efforts, however, have limitations by not being solely satellite-derived, by only studying certain regions of the oceans, by only examining one or two OCS parameters, and/or by only utilizing one or two machine learning models for the estimation, etc.

As such, a systematic study to evaluate various machine learning models on a global scale to see if any such model could be used to accurately and reliably predict all OCS parameters based solely on satellite-derived data is highly needed. To this end, this study is intended to make a preliminary attempt.

2. Data and Methods

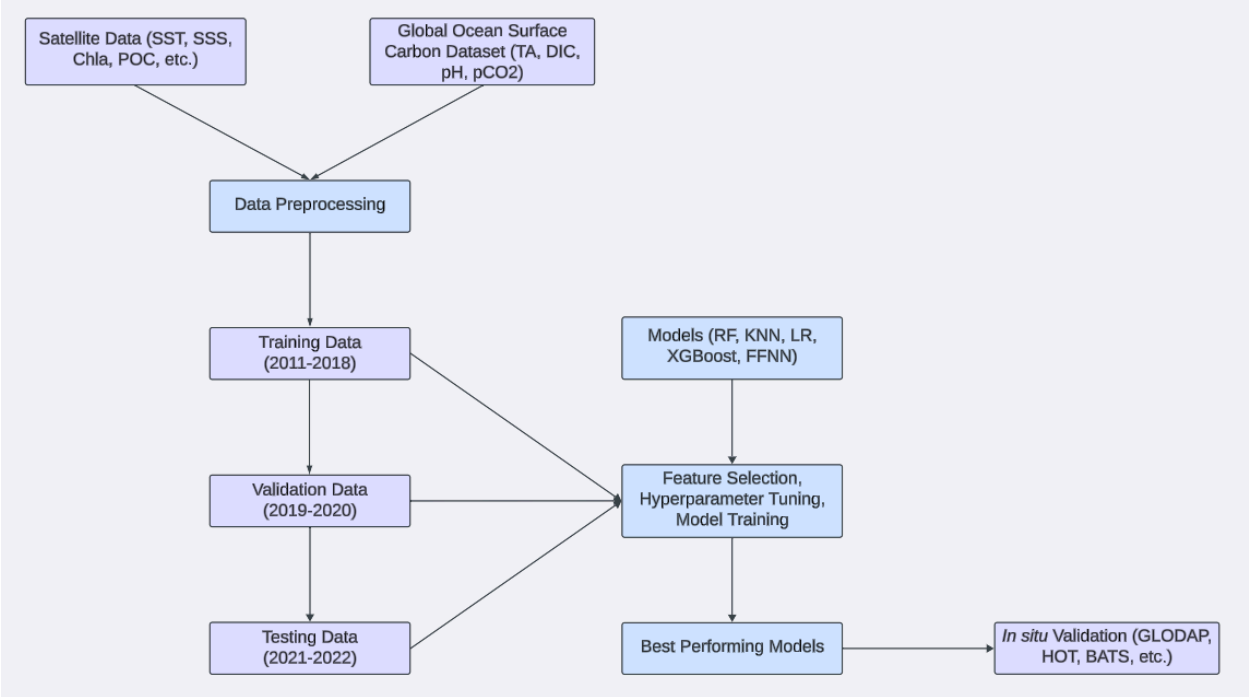

FIG. 1 Flowchart of methodology.

2.1 Target Data

Data for four OCS parameters, i.e. pH, DIC, TA, and pCO_2, were obtained from the Copernicus Marine Environment Monitoring Service (CMEMS) Global Ocean Surface Carbon Product (MULTIOBS_GLO_BIO_CARBON_SURFACE_REP_015_008) [17]. This dataset is a modeled, gridded product with a spatial resolution of 0.25° x 0.25°, based on *in situ* measurements. Due to the scarcity of satellite-matchup *in situ* observations, especially for pH [18, 26], modeled oceanic carbonate system data was used in place of *in situ* data to ensure a sufficient dataset for model training.

2.2 Satellite Data

The monthly and averaged satellite data from 2011-2024 were primarily obtained from the AQUA-MODIS (Moderate Resolution Imaging Spectroradiometer), METOP-ASCAT (Advanced Scatterometer aboard METOP-A, METOP-B, METOP-C), Aquarius/SAC-D, Soil Moisture Active Passive (SMAP), and Soil Moisture and Ocean Salinity (SMOS) missions.

These missions provided data for the following predictors: sea surface temperature (SST), chlorophyll-a (Chla), particulate inorganic carbon (PIC), particulate organic carbon (POC), sea surface salinity (SSS), zonal wind speed (u-wind), and meridional wind speed (v-wind). Specifically, SST, Chla, PIC, and POC data were obtained from AQUA-MODIS; SSS data were obtained from OISSS (Optimal Interpolation of Sea Surface Salinity) which aggregates observations from the Aquarius/SAC-D, SMAP, and SMOS missions [19]; u-wind and v-wind data were obtained from METOP-ASCAT, using data from METOP-A, METOP-B, and METOP-C. In cases where multiple METOP satellites were simultaneously operational, data from newer satellites were prioritized over older ones.

All data were resampled, using bilinear interpolation, to a spatial resolution of 0.25° x 0.25° to match up with the gridded target data. Additionally, the following transformations of satellite data were tested as potential predictors:

$$\text{cmon} = \cos\left(\frac{\text{month} \cdot \pi}{180}\right) \quad (1)$$

$$\text{smon} = \sin\left(\frac{\text{month} \cdot \pi}{180}\right) \quad (2)$$

$$\text{clon} = \cos\left(\frac{\text{longitude} \cdot \pi}{180}\right) \quad (3)$$

$$\text{slon} = \sin\left(\frac{\text{longitude} \cdot \pi}{180}\right) \quad (4)$$

$$\text{pic/poc} = \frac{\text{PIC}}{\text{POC}} \quad (5)$$

$$\text{logchla} = \log(\text{Chla}) \quad (6)$$

$$\text{windspeed} = \sqrt{\text{u-wind}^2 + \text{v-wind}^2} \quad (7)$$

$$\text{winddir} = \left(\frac{180}{\pi} \cdot \arctan\left(\frac{\text{u-wind}}{\text{v-wind}}\right) + 360\right) \% \, 360 \quad (8)$$

Transformations specified in equations (1), (2), (3), and (4) aim to capture the periodicity of month and longitude data [20]. The PIC:POC ratio specified in equation (5) has been considered an important term for modeling carbon cycling in the oceans [21, 22]. The log transformation applied to Chla in equation (6) aims to account for Chla's lognormal global distribution [23, 24]. Lastly, transformations applied to u-wind and v-wind in equations (7) and (8) aim to capture wind speed and directional effects more effectively.

2.3 *In Situ* Validation Data

In order to evaluate the robustness and accuracy of trained models in predicting OCS parameters based solely on satellite derived data, the following reported datasets (summarized in Table 1) were further utilized for *in situ* validation.

Table 1. *In situ* Validation Datasets used in this study.

Dataset Name	**Measured Parameters**	**Description**	**References**
HOT (Hawaiian Ocean Time Series)	TA, DIC	Long-term, high-quality dataset in the Pacific Ocean, located at 22.57° N, 158° W. HOT measurements for TA and DIC have an accuracy of < 2 μmol kg^{-1} and < ~4 μmol kg^{-1} respectively.	[16, 25]

BATS (Bermuda Atlantic Time Series)	TA, DIC	Long-term, high-quality dataset in the Atlantic Ocean, located at 32° N, 64° W. BATS measurements for TA and DIC have an accuracy of < 2 μmol kg^{-1}	[26, 27]
LDEO (Lamont Doherty Earth Observatory)	pCO_2	Global surface pCO_2 data from multiple sources, with large spatial distribution. The uncertainty of measurements is estimated to be ± 2.5 μatm	[28]
SOCAT (Surface Ocean CO_2 Atlas)	pCO_2	Global surface pCO_2 data from multiple sources, with large spatial distribution. Data with Quality Control (QC) flags of A, B, C, D, were chosen, with accuracy < 5 μatm.	[29]
GLODAPv2 (Global Ocean Data Analysis Project)	TA, DIC, pH	Global *in situ* TA, DIC, and pH data from multiple sources, with large spatial distribution. GLODAP measurements for TA, DIC, and pH have accuracies of < 4 μmol kg^{-1}, < 4 μmol kg^{-1}, and < 0.1–0.2 (depending on region) respectively.	[30]
SOCCOM (Southern Ocean Carbon and Climate Observations and Modeling)	pH	Autonomous float measurements of pH in the Southern Ocean, with a mean uncertainty of ± 0.019 for pH.	[31]

Data inclusion and exclusion criteria were applied based on previous studies [32, 33]. Specifically, *in situ* TA, DIC, and pH data at depths shallower than 30 m and in latitudes greater than 30° N, and at depths shallower than 20 m in latitudes lower than 30° N were selected for the *in situ* validation datasets. To remove outliers for surface pCO_2, the satellite-derived SST (T^{SST}) was first adjusted from the *in situ* measured temperature ($T^{in\ situ}$) using the following formula:

$$pCO_2^{SST} = pCO_2^{in\ situ} \times \exp\left(0.0433 \cdot \left(T^{SST} - T^{in\ situ}\right)\right) \quad (9)$$

and when the absolute difference between pCO_2^{SST} and $pCO_2^{in\ situ}$ is greater than 40 μatm, the surface pCO_2 data was regarded as an outlier and was excluded from the *in situ* validation datasets.

2.4 Model Development

Five machine learning models that have previously been shown to have relatively good performance in estimating OCS parameters, including Random Forest (RF), K-Nearest Neighbors (KNN), XGBoost (XGB), Feed-forward Neural Networks (FFNN), and Linear Regression (LR), were separately tested and compared in this study [16, 18, 34, 35]. Model development was done using Python, through the scikit-learn, TensorFlow, and XGBoost libraries.

The entire satellite-matchup dataset was split into a training set, validation set, and testing set. To prevent information leakage and reduce temporal autocorrelation while evaluating the feasibility real-time estimates of OCS parameters, the training, validation, and testing sets were split temporally, rather than randomly, as previously reported [36]. After splitting, the training set consisted of data from the years 2011-2018 (~40,000,000 data points), the validation set consisted of data from the years 2019-2020 (~10,000,000 data points), and the testing set consisted of data from the years 2021-2022 (~10,000,000 data points).

Of the five models tested for each parameter, the one with the lowest RMSE was chosen for further *in situ* validation.

2.4.1 Feature Selection

Feature selection was conducted to identify the most relevant predictors in estimating the OCS parameters in order to optimize model performance. In this study, two primary methods were used: feature importance analysis and correlation analysis.

The feature importance analysis was conducted using random forest. Briefly, a generic, random forest model was trained using the training set and feature importances were calculated based on the mean reduction of Mean Squared Error (MSE, equation (10)), which reflects the variance reduction achieved by each feature across all splits of the forest. Feature importances

were sorted, with the top 50% of features prioritized for inclusion within the final model, while features in the bottom 50% were excluded.

The correlation analysis was performed by calculating the pairwise Pearson Correlation Coefficient (R, equation (11)) between all features, providing insights on possible collinearity between predictors to avoid redundancy. When two features both had high feature importances but also a high correlation of $|R| > 0.8$ with one another, the feature with the highest feature importance was prioritized and the other was considered for exclusion.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \tag{10}$$

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{11}$$

Multiple experiments were then conducted through training on subsets of flagged features. Each experiment was trained on a random subset (10%) of the training data so as to reduce the computational load while still capturing the variability of the data. The experimental feature configuration with the lowest Root Mean Square Error (RMSE) when evaluated by the validation set was chosen.

2.4.2 Hyperparameter Tuning

Due to the large scale of the training set, a subset (10%) from the training set data was randomly chosen for hyperparameter tuning. Grid search was used to iterate over all hyperparameter configurations to determine the best performing configuration. Model configurations with the lowest RMSE when evaluated by the validation set were chosen for training on the entire training dataset, and eventual validation against *in situ* data.

2.4.3 Model Evaluation

Models were evaluated based on three metrics including coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Bias (MB), whose formulas are shown below. R^2 reflects the proportion of variance explained by the model; RMSE measures the average magnitude of prediction errors, specifically penalizing larger errors; and MB calculates the average distance between predicted and actual values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (12)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (13)$$

$$\text{Mean Bias} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \quad (14)$$

3. Results and Discussion

3.1 Evaluation of Model Performance

To find a solely satellite-based approach that can reliably predict the various OCS parameters, five machine learning models (RF, LR, KNN, XGBoost, and FFNN) were systematically evaluated and compared with respect to their performances in predicting the various OCS parameters (TA, DIC, pH, and pCO_2) on a global, regional, and monthly scale. Feature selection and hyperparameter tuning were conducted as outlined in Sections 2.4.1 and 2.4.2. For example, the final selected features for RF are listed in Table 2. The performance of each model for predicting each OCS parameter is summarized and compared in Table 3.

Table 2. Selected features for RF to predict the four oceanic carbonate system parameters (TA, DIC, pH, and pCO_2).

Parameter	**Features**
TA	Latitude, Longitude, SST, SSS, POC
DIC	Latitude, Longitude, SST, SSS, Wind Speed, Wind Direction
pH	Latitude, clon, slon, SST, SSS, Chla, Wind Speed, Wind Direction
pCO_2	Latitude, Longitude, SST, SSS, Chla, Wind Speed, Wind Direction

Table 3. Performance comparison of various models (RF, LR, KNN, XGBoost, and FFNN) in predicting each OCS parameter (TA, DIC, pH, and pCO_2) on the testing set, based on the satellite data.

Model	TA (μmol kg^{-1})		DIC (μmol kg^{-1})		pH		pCO_2 (μatm)	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
RF	0.97	11.81	0.97	13.63	0.75	0.02	0.74	17.74
LR	0.96	13.58	0.91	21.60	0.26	0.03	0.24	28.23
KNN	0.91	20.58	0.95	15.30	0.71	0.02	0.70	18.70
XGBoost	0.97	11.82	0.96	13.79	0.74	0.02	0.73	17.78
FFNN	0.94	17.40	0.93	19.20	-0.32	0.03	0.51	24.09

As shown in Table 3, among the five models tested in this study, the best performing model was RF, as evidenced by the highest R^2 values and lowest RMSE values across all OCS parameters. I speculated that the superior performance for RF is likely due to the fact that RF is essentially an ensemble approach, which utilizes multiple decision trees to capture complex patterns. XGBoost also noticeably performed well in predicting each of the four OCS parameters, second only to RF and with the R^2 and RMSE values substantially comparable to RF. KNN's performances were slightly lower than RF and XGBoost, but still achieved R^2 in a similar range as RF and XGBoost, i.e. $R^2 \geq 0.90$ in predicting TA and DIC, and $R^2 \geq 0.70$ in predicting pH and pCO_2. LR demonstrated relatively decent performances in predicting TA and DIC ($R^2 > 0.90$), but struggled in predicting pH and pCO_2, yielding low R^2 values of 0.26 and 0.24 respectively. The latter poor prediction performances are likely the result of the linear nature of LR, which fails to capture nonlinear relationships present between the features. FFNN also demonstrated decent performances in predicting TA and DIC ($R^2 > 0.90$), and also similarly struggled in predicting pH and pCO_2.

As further shown in Table 3, among the four OCS parameters investigated in this study, all five models can predict TA and DIC with decently good performances, achieving $R^2 > 0.90$, yet the performances in predicting pH and pCO_2 were all notably lower ($R^2 \leq 0.75$). For pH and pCO_2, RF achieved an R^2 of 0.75 and 0.74 respectively, a result notably aligning with the performance of past modeling efforts that produced similar $R^2 < 0.85$ [18, 34].

Because RF achieved the best prediction performances across all OCS parameters, RF was selected as the optimal model for the subsequent analyses and *in situ* validation.

3.2 Spatial and Temporal Analysis

RF was first evaluated to examine if there are any basin-scale differences in predicting the OCS parameters based on satellite data. For this purpose, the Atlantic, Pacific, Indian, Arctic, and Southern oceanic basins, which are specified in the RECCAP2 project [37] based on the boundaries defined by Fay and McKinley [38] (the five oceanic basins are illustrated in FIG. 2), were investigated, and the performances of RF in predicting each of the four OCS parameters in these five major oceanic basins were compared, with results summarized in Table 4.

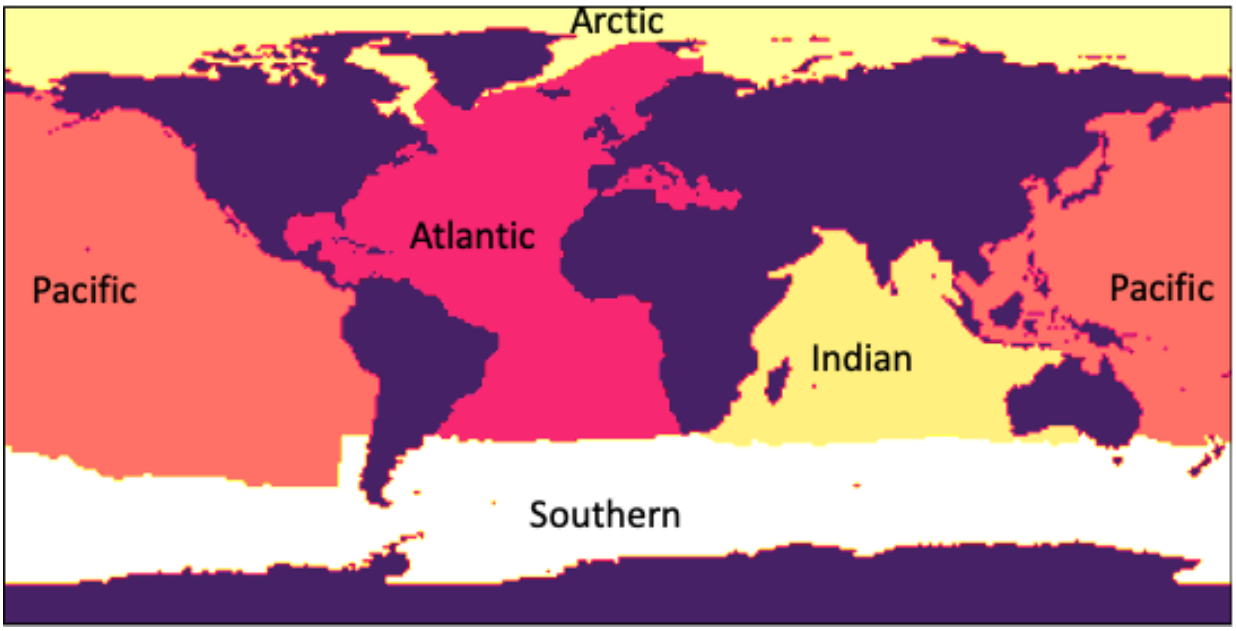

FIG. 2 Five oceanic basins (Atlantic, Pacific, Indian, Arctic, and Southern) examined in this study (adapted from [38]).

Table 4. RF performance comparison of the five oceanic basins (Atlantic, Pacific, Indian, Arctic, and Southern) in predicting each OCS parameter (TA, DIC, pH, and pCO_2) on the testing set, based on satellite data.

Region	TA (μmol kg^{-1})			DIC (μmol kg^{-1})		
	R^2	RMSE	n	R^2	RMSE	n
Atlantic	0.95	14.49	2,214,608	0.92	14.79	2,213,652
Pacific	0.96	10.51	4,769,475	0.94	13.72	4,768,592
Indian	0.95	12.08	1,396,544	0.95	13.47	1,396,298
Arctic	0.96	31.30	156,551	0.94	27.12	156,379
Southern	0.92	5.70	2,019,812	0.94	9.63	2,019,811
Global	0.97	11.81	10,601,256	0.97	13.63	10,598,925
Region	pH			pCO_2 (μatm)		
	R^2	RMSE	n	R^2	RMSE	n
Atlantic	0.69	0.02	2,213,652	0.70	17.11	2,214,608
Pacific	0.75	0.02	4,768,592	0.74	19.36	4,769,475
Indian	0.66	0.01	1,396,298	0.51	16.10	1,396,544
Arctic	0.73	0.03	156,379	0.77	21.00	156,551
Southern	0.50	0.02	2,019,811	0.40	14.79	2,019,812
Global	0.75	0.02	10,598,925	0.74	17.74	10,601,256

As shown in Table 4, there existed performance differences in RF across various oceanic basins. Notably, RF struggled the most in predicting OCS parameters in the Arctic Ocean when compared with other oceanic basins, as evidenced by the largest RMSEs across all parameters compared with other oceanic basins. I speculated that this may be related to the relatively small sample size of the Arctic Ocean data, which make up only < 2% of all testing data. In addition, the limitations of satellite-observations in this oceanic basin, as well as the Arctic's complex interactions between ice cover and temperature fluctuations as previously reported in the Arctic Ocean [39] may also contribute to the model's difficulty in producing accurate predictions. When RF was used to predict pH and pCO_2, the Southern Ocean yielded lowest R^2 values ($R^2 \leq 0.50$) compared with other oceanic basins, which may also be due to the complex sea ice-air interactions within the Southern Ocean [40].

RF was next evaluated across each month to examine if there are any temporal (e.g. monthly or seasonal) differences in predicting the OCS parameters based on satellite data, and the comparison results are shown in Table 5.

Table 5. RF performance comparison of the 12 months in predicting each OCS parameter (TA, DIC, pH, and pCO2) on the testing set, based on satellite data.

Month	TA (μmol kg^{-1})			DIC (μmol kg^{-1})			pH			pCO_2 (μatm)		
	R^2	RMSE	n	R^2	RMSE	n	R^2	RMSE	n	R^2	RMSE	n
January	0.97	10.44	953,996	0.97	12.31	953,842	0.73	0.02	953,842	0.73	16.64	953,996
February	0.98	9.41	974,443	0.97	11.87	974,285	0.76	0.02	974,285	0.74	17.18	974,443
March	0.98	10.12	979,303	0.97	12.04	979,134	0.73	0.02	979,134	0.69	17.76	979,303
April	0.98	10.72	815,060	0.97	13.29	814,897	0.67	0.02	814,897	0.65	18.85	815,060
May	0.98	11.20	805,060	0.96	14.01	804,883	0.76	0.02	804,883	0.75	18.27	805,060
June	0.98	12.38	757,177	0.95	15.20	756,954	0.82	0.02	756,954	0.80	18.51	757,177
July	0.97	14.65	788,056	0.95	15.99	787,782	0.85	0.02	787,782	0.83	17.82	788,056
August	0.98	13.26	868,261	0.97	14.42	868,005	0.80	0.02	868,005	0.80	18.06	868,261
September	0.97	14.73	916,138	0.97	14.36	915,887	0.74	0.02	915,887	0.75	17.15	916,138
October	0.97	12.14	935,300	0.97	13.32	935,118	0.65	0.02	935,118	0.63	16.71	935,300
November	0.97	11.32	905,848	0.97	13.31	905,683	0.57	0.02	905,683	0.55	17.08	905,848
December	0.97	10.75	902,614	0.96	13.84	902,455	0.60	0.02	902,455	0.60	19.15	902,614

As shown by the month-by-month comparison results in Table 5, when RF was used to predict the OCS parameters based on the satellite-derived data, there also exist performance differences across different months. For TA and DIC, predictions made during Northern Hemisphere summer months, i.e. June, July, August, September, are notably characterized by relatively higher RMSEs compared with other months. I speculated that this may be caused by the influences of the Arctic, where substantial ice melting occurs typically during those months. There is a noticeable drop in the amount of data during those months, which may also influence the RMSE values.

3.3 *In Situ* Validation and Comparison to Existing Models

In order to further assess the model robustness and accuracy, RF was evaluated against several publicly available *in situ* datasets (i.e. GLODAP, HOT, BATS, SOCCOM, SOCAT, and LDEO, as described above in Table 1) for *in situ* validation. Table 6 summarizes the validation result, which was also compared with several existing studies on the same *in situ* datasets.

As shown in Table 6, the solely satellite-based RF model in this study was capable of predicting at least some of the OCS parameters at high accuracies. Among the six *in situ* datasets (i.e. GLODAP, HOT, BATS, SOCCOM, SOCAT, and LDEO) against which the RF model was validated, the validation over GLODAP generated the best validation results for TA, DIC, and pH, and the scatter plots depicting the relationship of the model-predicted value versus actual values are further shown in FIG. 2(a)-(c) respectively.

Table 6. Performance comparison of the RF model validated based on *in situ* datasets (GLODAP, HOT, BATS, SOCCOM, SOCAT, and LDEO) with several previously reported *in situ*-based models (GRaCER, LIAR + FFNN, NNGv2, and SOMFFNN).

		TA (μmol kg^{-1})			DIC (μmol kg^{-1})			pH		pCO_2 (μatm)	
		GLODAP	HOT	BATS	GLODAP	HOT	BATS	GLODAP	SOCCOM	SOCAT	LDEO
R^2	this study	0.87	0.6	0.27	0.95	0.32	0.60	0.68	-0.06	0.54	0.70
	GRaCER	0.91	0.58	0.13	0.93	0.77	0.76	0.67	0.05	0.82	0.45
	LIAR + FFNN	0.91	0.6	0.21				0.67	-0.05	0.78	0.44
	NNGv2	0.93	0.75	-0.1	0.82	0.71	0.38				
	SOMFFNN									0.82	0.49
RMSE	this study	11.7	10.3	10.8	16.8	11.4	10.8	0.025	0.043	34.3	18.6
	GRaCER	17.5	9.5	10.1	16.3	8.7	9.1	0.024	0.036	11.1	19.9
	LIAR + FFNN	18.0	8.8	8.8				0.023	0.037	13.1	19.6
	NNGv2	16.2	6.7	10.4	23.1	9.5	15.2				
	SOMFFNN									11.7	21.4
Bias	this study	-1.4	-4.0	-2.7	-1.7	3.4	-4.3	0.002	0.032	-12.4	1.96
	GRaCER	0.5	-2.1	2.6	0.5	-1.0	0.4	-0.001	0.009	-0.4	0.1
	LIAR + FFNN	0.3	-3.1	0.2				0.001	0.013	0.5	0.6
	NNGv2	1.2	-3.2	4.3	2.3	2.2	-0.4				
	SOMFFNN									0.4	0.4

Note: The *in situ* datasets (GLODAP, HOT, BATS, SOCCOM, SOCAT, and LDEO) were described in Table 1. The performance data for all of the existing *in situ*-based models (GRaCER, LIAR + FFNN, NNGv2, and SOMFFNN) were taken from Gregor and Gruber (2020) [4].

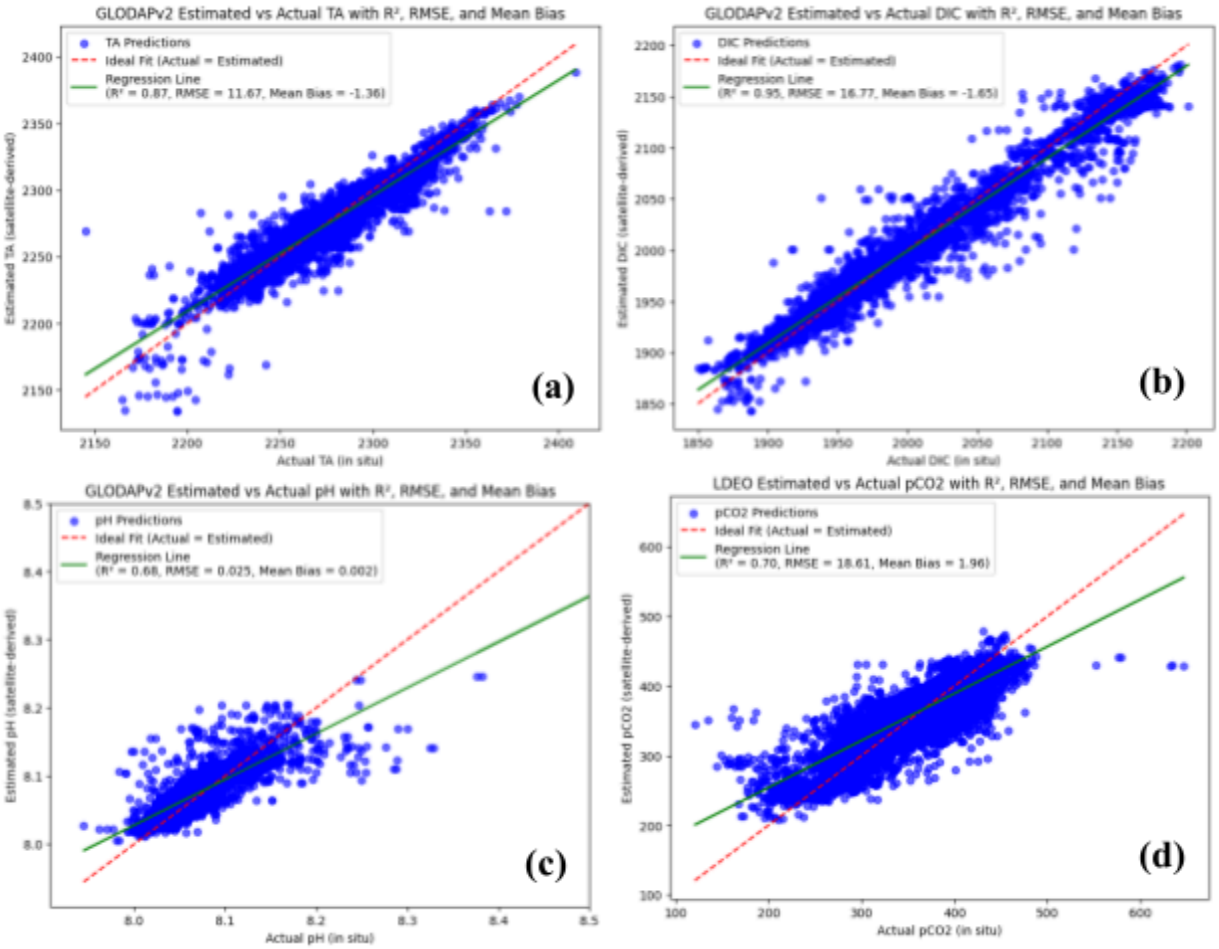

FIG. 2 Scatter plots for the model-predicted values vs actual values for the different oceanic carbonate system parameters, including TA (a); DIC (b); pH (c); and pCO2 (d). Among these, (a)-(c) were based on the *in situ* validation dataset GLODAP, and (d) was based on the *in situ* validation dataset LDEO.

Regarding TA, the RF model performed well when evaluated against the *in situ* validation dataset GLODAP, as illustrated by the matching of the model-based TA predictions with the actual *in situ*-obtained TA values in FIG. 2(a). Although the RF model yielded a slightly lower R^2 value (0.87) in predicting TA compared with the reported GRaCER study (R^2 = 0.91), LIAR + FFNN (R^2 = 0.91), and NNGv2 (R^2 = 0.93), the model had a significantly lower RMSE value (11.7 μmol kg^{-1}) than these three reported models (RMSEs ranges 16.2-18.0 μmol kg^{-1}).

Based on the validation of DIC over the *in situ* dataset GLODAP, it can be observed that the solely satellite-based RF model could also predict DIC very well (R^2 = 0.95, see Table 6), and that model-based DIC predictions matched very well with the actual *in situ*-obtained DIC values (see FIG. 2(b)). Notably the performance of the RF model is better than the reported GRaCER study (R^2 = 0.93) and NNGv2 study (R^2 = 0.82).

As also illustrated by FIG. 2(c) and shown in Table 6, the RF model may also predict pH with a decent accuracy, which yielded a R^2 value of 0.68 when validated also against the *in situ* validation dataset GLODAP. Although this R^2 value is not as high as that for the validation of TA (R^2 = 0.87) and DIC (R^2 = 0.95), it is notably higher than the reported GRaCER study (R^2 = 0.67) and LIAR+FFNN study (R^2 = 0.67).

It is further noteworthy that based on validation over the *in situ* dataset LDEO, the solely satellite data-based RF model in this study yielded a significantly better R^2 value (0.70) in predicting pCO_2 (see FIG. 2(d)) compared with the reported GRaCER (R^2 = 0.45) and LIAR+FFNN (R^2 = 0.44) (see Table 6), and consistently yielded an RMSE (RMSE = 18.6 μatm) that was lower than GRaCER (RMSE = 19.9 μatm) and LIAR+FFNN (RMSE = 19.6 μatm).

These above *in situ* validation results may indicate the RF model as a potentially useful machine learning model capable of accurately predicting these oceanic carbonate system parameters based solely on satellite data and, at the very least, demonstrate the feasibility of monitoring of the oceanic carbonate system based solely on satellites.

It should be noted, however, the RF model in this study seemed to produce biases that were slightly larger in magnitude compared to other reported *in situ*-based methods (please see the "Biases" rows for each OCS parameter in Table 6). I speculated that this may be related to the solely satellite-based nature of this study, and may as well be related to the fact that the model was trained on monthly gridded data rather than discrete *in situ* observations.

Additionally there seem to exist performance differences among different *in situ* validation datasets. Regarding TA, for example, while the validation over the *in situ* datasets GLODAP and HOT yielded R^2 of 0.87 and 0.60 respectively for the RF model, the validation over BATS only yielded a surprisingly low R^2 of 0.27. Regarding DIC, the validation over GLODAP and BATS yielded R^2 of 0.95 and 0.60 respectively for the RF model, whereas the validation over HOT only yielded a surprisingly low R^2 of 0.32. Thus each of the HOT and BATS *in situ* validation datasets produced strikingly different performances for the

under-validation RF model in this study. This may reflect some systematic biases existing in these different *in situ* datasets. Therefore, there is a need for a follow-up validation study over additional *in situ* datasets.

4. Conclusion and Discussion

This study proposes a solely satellite-based approach to monitoring the oceanic carbonate system. Through testing of various hyperparameter and feature combinations, Random Forest (RF) was identified, among the five machine learning models (RF, LR, KNN, XGBoost, and FFNN) that were tested, to be the best performing model in predicting oceanic carbonate system parameters (i.e. TA, DIC, pH, and pCO_2) based solely on satellite-derived data. The spatial and temporal analysis further revealed certain limitations, particularly in regions such as the Arctic and Southern Oceans. When further validated against *in situ* datasets, RF demonstrated a decent predictive capability, which is notably comparable to that of current *in situ*-based modeling efforts.

Through this systematic effort, this study has demonstrated the feasibility of monitoring the OCS based solely on satellite data and has identified RF as a useful machine learning model for estimating these OCS parameters at good accuracy, potentially decreasing our reliance on expensive and environmentally harmful in situ sampling. However, there are certain caveats to take note of.

Firstly, the reliance on modeled carbonate data rather than *in situ* measurements for target data during the model development stage may introduce potential limitations. Since modeled data itself is an approximation based on *in situ* data, it may lack some of the fine-scale variability observed in *in situ* data. Especially with pH, which is rarely measured, and instead mathematically modeled through CO2SYS, it's difficult to fully validate modeled products.

Secondly, since the model was trained on monthly-averaged data, it may not capture short-term variability effectively. However, future studies incorporating data with higher temporal resolution (e.g. weekly, daily, or even hourly) may advance higher-resolution monitoring efforts of the oceanic carbonate system.

Thirdly, the temporal split of training, testing, and validation sets represents another potential limitation. While it was essential to avoid information leakage and reduce temporal autocorrelation so as to prevent overoptimistic estimates of model performance, such a split may impact the model's ability to learn the full range of seasonal variability. The training dataset was

only 8 years long, which may have limited the model from learning multi-year trends on the OCS parameters. Such a split may also influence higher errors in regions like the Arctic Ocean, which is characterized by a lower data count. Though performance on *in situ* datasets were comparable to that of current *in situ*-based modeling efforts, future work could benefit from using spatiotemporal cross-validation or an expanded temporal range.

Lastly, it shall be noted that the model development can be further optimized. Future studies may incorporate various other models and feature combinations to potentially improve prediction accuracy and capture more subtle trends within the data.

In light of these above caveats, a future more in-depth research is needed to further address the limitations, and to verify the findings, of this present study. Once verified and implemented, this machine learning and solely satellite data-based approach shall provide a green, cost-effective, and high spatiotemporal-resolution approach to supplement or even replace the current *in situ*-based OCS monitoring practices.

References:

1. Feely, R. A., et al. Impact of Anthropogenic CO_2 on the $CaCO_3$ System in the Oceans. *Science*, 305, 362–366, 2004.
2. Sabine, C. L., et al. The Oceanic Sink for Anthropogenic CO_2, *Science*, 305, 367–371, 2004.
3. Doney, S. C., et al. Ocean acidification: the other CO_2 problem. *Annu. Rev. Mar. Sci.* 1, 169–192 (2009).
4. Gattuso, J.-P. *et al*. Contrasting futures for ocean and society from different anthropogenic CO_2 emissions scenarios. *Science* 349, (2015)
5. Anthony, K. R. N. et al.. (2008). Ocean acidification causes bleaching and productivity loss in coral reef builders. *Proceedings of the National Academy of Sciences*. 105 (45): 17442–17446.
6. Riebesell, Ulf, et al. (2000). Reduced calcification of marine plankton in response to increased atmospheric CO_2. *Nature*. 407 (6802): 364–367.
7. Jiang, Li-Qing; et al. (2019). Surface ocean pH and buffer capacity: past, present and future. *Scientific Reports*. 9 (1): 18624.
8. Chan NCS, Connolly SR. Glob Chang Biol. 2013 Jan;19(1):282-90.

9. Mollica N.R., Guo W, et al. (2018) *Proceedings of the National Academy of Sciences*, 115 (8) 1754-1759.
10. Bushinsky, S.M., Takeshita, Y. & Williams, N.L. Observing Changes in Ocean Carbonate Chemistry: Our Autonomous Future. *Curr Clim Change Rep* **5**, 207–220 (2019).
11. Fernandes, P.G., Stevenson, P. and Brierley, A.S., (2002). AUVs as research vessels: the pros and cons. ICES CM2002/J:02, 11 pp
12. Williams NL, Juranek LW, Feely RA, Russell JL, Johnson KS, Hales B. Assessment of the carbonate chemistry seasonal cycles in the Southern Ocean from persistent observational platforms. J Geophys Res Ocean. 2018;123:4833–52.
13. Gray AR, Johnson KS, Bushinsky SM, Riser SC, Russell JL, Talley LD, et al. Autonomous biogeochemical floats detect significant carbon dioxide outgassing in the high-latitude Southern Ocean. Geophys Res Lett. 2018;45:9049–57.
14. V. Sarma, T. Saino, K. Sasaoka, *et al.* Basin-scale pCO_2 distribution using satellite sea surface temperature, Chl *a*, and climatological salinity in the North Pacific in spring and summer. Global Biogeochem Cy, 20 (2006).
15. K. Lee, L. T. Tong, F. J. Millero, C. L. Sabine, A. G. Dickson, C. Goyet, et al., "Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans", *Geophys. Res. Lett.*, vol. 33, no. 19, Oct. 2006.
16. Gregor, Luke, and Nicolas Gruber. "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 Discussions 2020 (2020): 1-42.
17. Chau, T. T., Marion Gehlen, and Frédéric Chevallier. "Global Ocean Surface Carbon Product MULTIOBS_GLO_BIO_CARBON_SURFACE_REP_015_008." EU Copernicus Marine Service Information 3.0 (2020): 15.
18. Jiang, Zhiting, et al. "Remote sensing of global sea surface pH based on massive underway data and machine learning." Remote Sensing 14.10 (2022): 2366.
19. Melnichenko, Oleg, et al. "Multi-Mission Sea Surface Salinity Optimum Interpolation (OISSS) Analysis Version 2.0." PODAAC: Pasadena, CA, USA (2023).
20. Zeng, J., et al. "A global surface ocean fCO2 climatology based on a feed-forward neural network." Journal of Atmospheric and Oceanic Technology 31.8 (2014): 1838-1849.

21. Ridgwell, A., et al. "From laboratory manipulations to Earth system models: scaling calcification impacts of ocean acidification." Biogeosciences 6.11 (2009): 2611-2623.
22. Findlay, Helen S., Piero Calosi, and Katharine Crawfurd. "Determinants of the PIC: POC response in the coccolithophore Emiliania huxleyi under future ocean acidification scenarios." Limnology and Oceanography 56.3 (2011): 1168-1178.
23. Campbell, Janet W. "The lognormal distribution as a model for bio‐optical variability in the sea." Journal of Geophysical Research: Oceans 100.C7 (1995): 13237-13254.
24. Vantrepotte, Vincent, and Frederic Mélin. "Temporal variability of 10-year global SeaWiFS time-series of phytoplankton chlorophyll a concentration." ICES Journal of Marine Science 66.7 (2009): 1547-1556.
25. Dore, John E., et al. "Physical and biogeochemical modulation of ocean acidification in the central North Pacific." Proceedings of the National Academy of Sciences 106.30 (2009): 12235-12240.
26. Bates, Nicholas R., et al. "A time-series view of changing surface ocean chemistry due to ocean uptake of anthropogenic CO_2 and ocean acidification." Oceanography 27.1 (2014): 126-141.
27. Bates, Nicholas R., and Rodney J. Johnson. "Forty years of ocean acidification observations (1983–2023) in the Sargasso Sea at the Bermuda Atlantic Time-series Study site." Frontiers in Marine Science 10 (2023): 1289931.
28. Takahashi, Taro, Stewart C. Sutherland, and Alex Kozyr. Global ocean surface water partial pressure of CO2 database: Measurements performed during 1957–2018 (version 2018). NOAA/NCEI/OCADS NDP-088 (V2018) Rep., 25 pp., https://www. ncei. noaa. gov/access/ocean-carbon-data-system/oceans/LDEO_Underway_Database/NDP-088_V2018. Pdf (last access: October 2023), 2019.
29. Bakker, Dorothee C., et al. "Surface ocean CO2 atlas database version 2022 (SOCATv2022)(ncei accession 0253659)." (2022).
30. Lauvset, Siv K., et al. "GLODAPv2. 2022: the latest version of the global interior ocean biogeochemical data product." Earth System Science Data 14.12 (2022): 5543-5572.
31. Johnson, Kenneth S., et al. "Biogeochemical sensor performance in the SOCCOM profiling float array." Journal of Geophysical Research: Oceans 122.8 (2017): 6416-6436.

32. Lee, Kitack, et al. "Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans." Geophysical research letters 33.19 (2006).
33. Goddijn-Murphy, L. M., et al. "The OceanFlux Greenhouse Gases methodology for deriving a sea surface climatology of CO 2 fugacity in support of air–sea gas flux studies." Ocean Science 11.4 (2015): 519-541.
34. Chau, Thi-Tuyet-Trang, et al. "CMEMS-LSCE: a global, 0.25∘, monthly reconstruction of the surface ocean carbonate system." Earth System Science Data 16.1 (2024): 121-160.
35. Broullón, Daniel, et al. "A global monthly climatology of total alkalinity: a neural network approach." Earth System Science Data 11.3 (2019): 1109-1127.
36. Botache, Diego, et al. "Unraveling the Complexity of Splitting Sequential Data: Tackling Challenges in Video and Time Series Analysis." arXiv preprint arXiv:2307.14294 (2023).
37. Terhaar, Jens, et al. "Assessment of global ocean biogeochemistry models for ocean carbon sink estimates in RECCAP2 and recommendations for future studies." Journal of Advances in Modeling Earth Systems 16.3 (2024): e2023MS003840.
38. Fay, A. R., and G. A. McKinley. "Global open-ocean biomes: mean and temporal variability." Earth System Science Data 6.2 (2014): 273-284.
39. Deng, Jiechun, and Aiguo Dai. "Sea ice–air interactions amplify multidecadal variability in the North Atlantic and Arctic region." Nature Communications 13.1 (2022): 2100.
40. Ayres, Holly C., et al. "The coupled atmosphere–ocean response to Antarctic sea ice loss." Journal of Climate 35.14 (2022): 4665-4685.