https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

Identifying Drivers of Sea Surface pCO_2 via the Lag-Convergent Cross Mapping Model

HuishengWu[1], Min Liu[1], Jinping Liu[2], Xiaochun Ling[3], Yang Wang[1], Zhuang Li[1], Miaomiao Qiu[1], Xiaoke Cheng[1], and Siyuan Ju[1]

[1]College of Oceanography and Spatial Information, China University of Petroleum, Qingdao, 266580, China
[2]Shengli Branch of Sinopec Geophysical Co., Ltd, Dongying 257086, China
[3]Shandong Provincial Institute of Land Surveying and Mapping, Jinan, Shandong 250102, China

Correspondence to: Huisheng Wu (wuhuisheng@upc.edu.cn)

Abstract. In the context of global climate change, analyzing the driving factors of key ocean carbon cycle parameters is crucial for accurately quantifying ocean carbon sink capacity. The Convergent Cross-Mapping (CCM) method provides an effective approach for causal inference in nonlinear systems, yet it cannot characterize the ubiquitous time-lag effects within ocean carbon cycle systems, which readily results in underestimated causal strength and misidentified causal relationships. To address this limitation, this study incorporates the causal time-lag parameter (*Lc*) into the CCM framework as an optimized variable, and proposes a Lag-Convergent Cross Mapping (L-CCM) approach that accounts for optimal causal time lags, thereby constructing a time-lag embedded causal inference model. Taking the subtropical Northwest Pacific as the study domain, this study combines monthly causal screening and daily quantitative time-lag analysis to identify 18 potential drivers of sea surface pCO_2. Results show that L-CCM detects optimal causal time lags, under which the average causal strength of all drivers increased by 10.32 %. Strong causal drivers of sea surface pCO_2 include sea surface temperature (SST), sea surface salinity (SSS), chlorophyll a concentration (Chl), pH, etc., whereas weak causal factors cover surface zonal and meridional currents (*U*o, *V*o), zonal and meridional geostrophic currents (*U*gos, *V*gos). This method effectively reduces underestimation and misidentification of causal links, offers a reliable framework for ocean carbon cycle causal analysis, and holds promise for wider applications.

1 Introduction

Against the backdrop of global climate change, the ocean, as one of the largest active carbon sinks, plays a critical buffering role in the climate system by regulating atmospheric CO_2 concentrations (Tjiputra et al., 2025). Air-sea carbon dioxide exchange constitutes a core component of ocean carbon cycling. Sea surface partial pressure of carbon dioxide (pCO_2) describes the carbon dioxide concentration when gas exchange between seawater and the atmosphere reaches dynamic equilibrium (Zhong et al., 2022). It acts as a core indicator to characterize air-sea CO_2 exchange (Wang et al., 2014), and serves as an essential basis for exploring ocean carbon cycles and global climate change. Systematic exploration of this

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

parameter and its potential drivers facilitates the interpretation of carbon cycling mechanisms, and provides scientific references for addressing climate change issues.

Ocean carbon sinks are jointly modulated by coupled physical, chemical and biological processes, which drive the evolution of ocean carbon cycling under combined effects of multiple factors (Katavouta and Williams, 2021). Previous studies revealed that seawater temperature, salinity and other factors can affect sea surface pCO_2 (López-Puertas et al., 2025). Takahashi et al. (Takahashi et al., 2009), based on observations of global sea surface pCO_2, identified temperature, salinity, wind speed, sea ice cover, and biological productivity as the primary factors influencing pCO_2. Tozawa et al. (Tozawa et al., 2024) found that in the Pacific sector of the Arctic Ocean, biological activity and temperature variations exert significant impacts on pCO_2. Although statistical correlation analysis and machine learning inversion improved the fitting and prediction performance of sea surface pCO_2 to a certain extent (Chen et al., 2019), existing studies merely select influencing factors based on empirical correlation. Such methods fail to deliver causal evidence concerning interaction directions, influence pathways and temporal characteristics between driving factors and response variables. Restricted by correlation-based variable screening, the constructed models suffer from weak generalization capability and inconsistent accuracy among diverse marine regions, which is fundamentally attributed to the absence of causal mechanism support.

Dynamic system theory provides a powerful tool for identifying causal relationships within complex Earth systems.The Convergent Cross Mapping (CCM) method proposed by Sugihara et al. reconstructs the shadow manifold based on Takens' Embedding Theorem in dynamical systems to identify causal relationships, demonstrating robust causal detection capabilities in nonlinear, bidirectionally coupled, and chaotic systems (Sugihara et al., 2012). CCM has demonstrated outstanding advantages and gained extensive applications in ecology, climatology, medical science and other fields (Randall and van Woesik, 2017; Zou et al., 2021; Zhou et al., 2025). However, mainstream CCM implementations predominantly rely on in-phase cross-mapping to evaluate causal strength, overlooking the time lag with which driving factors influence response variables—a phenomenon referred to as causal time lag. Most models adopt a default of zero or extremely short time lags, assuming that driving and response variables act simultaneously, and thus fail to systematically characterize the temporal sequence of causal effects. This implicit assumption is inconsistent with the time-lag features widely present in complex system dynamics, particularly in ocean carbon sequestration processes. Existing studies demonstrated that both rapid responses (e.g., the reduction in pCO_2 driven by enhanced solubility under cooling conditions) (Takahashi et al., 2002) and slow responses (e.g., those associated with the biological pump) (Stange et al., 2017) introduce notable time lags between drivers and responses. If such process-related time lags are not explicitly incorporated into the analytical framework, they may lead to underestimation of the genuine causal strength or misinterpretation of causal directionality. This methodological limitation is not restricted to ocean carbon sink research, but also broadly hinders in-depth understanding of the underlying mechanisms governing numerous Earth system processes, including urban heat islands, epidemic transmission, and ecosystem dynamics.

Accordingly, this study explicitly introduces the causal time-lag parameter (Lc) into the complete CCM framework to quantify the effective time lag of driving factors acting on response variables. A novel causal identification method named

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

Lag-Convergent Cross Mapping (L-CCM) is established to comprehensively characterize causal strength and optimal lag scale. It is applied to identify causal relationships between sea surface pCO_2 and its potential drivers, verifying and demonstrating the improved causal identification capacity of this method. It provides methodological support and empirical evidence for shifting from correlation-based screening to causal constraints in subsequent ocean carbon sink inversion, overcomes the limitations of conventional models, and greatly enhances the accuracy and mechanistic interpretability of identifying drivers of ocean carbon cycling.

2 Materials and Methods

2.1 Study Area and Data Processing

2.1.1 Overview of the Study Area

We focused on the subtropical waters of the northwestern Pacific (28° N–40° N, 140° E–180° E). As one of the world's major ocean carbon sinks(Chen et al., 2025), this region exhibits substantial air-sea CO_2 exchange fluxes and is shaped by complex ocean dynamic processes, making it an ideal study area for identifying causal relationships among key components and driving factors of the ocean carbon sink.

We selected sea surface partial pressure of carbon dioxide (pCO_2) as the core parameter characterizing the ocean carbon sink, and identified its potential driving factors, including physical, chemical, and biological variables: sea surface temperature (SST), sea surface salinity (SSS), *p*H, chlorophyll a concentration (Chl), nitrate concentration (NO_3^-), dissolved oxygen concentration (O_2), total dissolved inorganic carbon (TCO_2), total alkalinity (TALK), air-sea CO_2 flux ($fgCO_2$), silicate concentration (Si), phosphate concentration (PO_4^{3-}), mixed layer depth (MLOTST), surface zonal and meridional currents (*U*o, *V*o), zonal and meridional geostrophic currents (*U*gos, *V*gos), and zonal and meridional wind speeds (*U*wind, *V*wind). These variables cover key ocean carbon sink processes, including the solubility pump, biological pump, and physical transport, thus providing comprehensive data support for identifying multifactorial causal relationships within a nonlinear coupled system.

2.1.2 Data Sources and Processing

We classified the data into two categories, which were respectively used for the preliminary diagnosis of macroscale causal relationships and the refined identification of causal time lags.

(a) Monthly-scale data

This section adopted CCM-based causal inference models to preliminarily diagnose the causal relationships between sea surface pCO_2 and its potential drivers. The sea surface pCO_2 observations were retrieved from the Surface Ocean CO_2 Atlas (SOCAT) (Bakker et al., 2016). This dataset is an authoritative compilation of global observed sea surface CO_2 partial pressure (pCO_2) data. It aggregates in-situ observations from platforms such as research vessels, merchant ships, and buoys worldwide, and has undergone rigorous quality control and standardization, providing reliable surface CO_2 observation data for this study; Satellite observation data were sourced from the NASA Ocean Color Level-3 products (NASA ocean color, 2026); numerical model data were sourced from the Copernicus Marine Environment Monitoring Service (CMEMS) Level-4 reanalysis products. These open-access datasets provide sufficient spatiotemporal coverage to meet the input requirements of the causal identification model. The data spanned from 2000 to 2020 at a monthly resolution. Following outlier removal, missing value processing, and spatiotemporal matching, 248 valid data points were obtained for each potential driver.

(b) Daily-scale Data

To finely resolve the causal time-lag relationships between potential drivers and sea surface $p CO_2$, we selected the mooring buoy station of the Kuroshio Extension Observatory (KEO, 32.3° N, 144.6° E) (Kuroshio extension observatory | ocean climate stations, 2026) within the study domain as the core site to conduct daily-scale causal time-lag analysis. KEO provides long-term, continuous time-series observational data on sea surface $p CO_2$ and related parameters from 2004 to the present. With the KEO site as the core location, priority was given to its five daily observational datasets—SST, SSS, Chl, O_2, and *p*H—which exhibit excellent continuity and have undergone strict quality control. To ensure comprehensive coverage of other potential drivers, the remaining variables were extracted from CMEMS regional grid products, which share a similar dynamic environment with the KEO site and provide stable daily-frequency data. Following preprocessing, 275 valid data points were obtained for each variable. This approach ensures consistency between the study subject and background oceanographic processes, while providing sufficient temporal resolution for causal inference models that account for time lags, within the constraints of data availability.

2.2 Research Methods

Traditional CCM ignores causal time lags between variables and fails to match the lagged response characteristics of multiple processes in ocean carbon systems. In this study, we introduced the causal lag parameter Lc to construct the L-CCM model. The model optimized state space reconstruction by matching the optimal response lag to achieve more accurate quantification of nonlinear causality. The implementation procedures of the model were outlined as follows:

2.2.1 Definition of the Lc Parameter

The causal time-lag parameter Lc (where $Lc \geq 0$) is defined as the time delay with which the driving factor X exerts its influence on the response variable Y. Within the cross-mapping framework, if variable X has a causal effect on Y with a time delay of Lc units, the state of Y at time $t + Lc$ will contain the information of X at time t. Consequently, the time-lagged embedded manifold M_Y^{Lc} can be used to estimate the state of X at time t.

We incorporated the causal time-lag parameter Lc into the CCM workflow and defined the cross-mapping prediction value at time t corresponding to Lc as: $\widehat{X}(t)|M_Y(t+\mathrm{L_c})$, where Lc denotes the causal time-lag parameter; when $Lc = 0$, this corresponds to the instantaneous causality assumption.

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

2.2.2 Design of the Global Scan Algorithm for *Lc*

To accurately identify the time-lag characteristics of causal relationships between variables, we developed a global scanning algorithm for causal time lags. By systematically traversing and quantitatively analyzing potential causal time-lag intervals, this algorithm enables quantitative identification of the optimal causal time lag. The algorithm takes the quantitative metric of causal strength as its core criterion. Within a reasonable time-lag search range, it calculates the causal time-lag parameter Lc point-by-point across all candidate values to obtain the causal strength $\rho(Lc)$ under different lags, and identifies the Lc that maximizes $\rho(Lc)$ as the optimal causal time lag Lc,opt.

Based on the physical mechanisms underlying ocean carbon sink processes and existing literature, the effects of potential drivers on sea surface $p\mathrm{CO_2}$ typically unfold over a finite timescale, ranging from several hours to tens of days (DeGrandpre et al., 2004; DeVries, 2022; Liu et al., 2024). We designed preliminary experiments to test and determine the maximum reasonable time lag. The results showed that the ρ value generally peaked and then gradually declined as Lc increased within a one-month window. Therefore, the scanning range adopted in this study was $Lc = 0, 1, 2, \ldots, 30$ (units: days), with the scanning period extended to 40 days for certain variables.

This procedure is executed point by point over the predefined lag grid $Lc \in [0, Lc, \mathrm{max}]$ to iterate across all candidate Lc values. It then derives the curve $\rho(Lc)$, which characterizes the variation in causal strength with time lag, and selects the Lc that maximizes $\rho(Lc)$ as the optimal causal time lag: $Lc, \mathrm{opt} = \mathrm{argmax}\,\rho(Lc)$. If multiple Lc values yield comparable peak $\rho(Lc)$, a comprehensive evaluation is conducted by integrating physical mechanisms and model convergence characteristics.

2.2.3 L-CCM Causal Inference Procedure

Let the original time series be:

$$X = \{X(1), X(2), \ldots, X(L)\}\,, \tag{1}$$

$$Y = \{Y(1), Y(2), \ldots, Y(L)\}\,, \tag{2}$$

where L denotes the total length of the time series (in days).

For a specified time lag Lc, construct a pair of lagged sequences,

$$X^{(Lc)} = \{X(1), X(2), \ldots, X(\mathrm{L} - L_c)\}\,, \tag{3}$$

$$Y^{(Lc)} = \{Y(1 + L_c), Y(2 + L_c), \ldots, Y(L)\}\,, \tag{4}$$

such that $X^{(Lc)}$ and $Y^{(Lc)}$ are temporally aligned, and $Y^{(Lc)}$ lags behind $X^{(Lc)}$ by Lc days. The effective sequence length is $L - Lc$. For the preprocessed time series, the embedding vector of $\mathrm{Y}(t + Lc)$ on the shadow manifold of $Y^{(Lc)}$ is defined as:

$$\mathrm{y}(\mathrm{t} + L_c) = \left(Y(t + L_c), Y(t + L_c - \tau), Y(t + L_c - 2\tau), \ldots, Y(t + L_c - (E - 1)\tau)\right)\,, \tag{5}$$

where $t \in [E+Lc, L]$.

Under time-delayed alignment, the embedding vector of X at time t is:

$$x(t) = (X(t), X(t - \tau), X(t - 2\tau), \ldots, X(t - (E - 1)\tau))\,, \tag{6}$$

where $t \in [E, L - Lc]$.

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

Construct the time-lagged state vector matrices $\mathrm{M}_{\mathrm{Y}}^{\mathrm{Lc}}$、 $\mathrm{M}_{\mathrm{X}}^{\mathrm{Lc}}$,

$$M_Y^{Lc} = \begin{bmatrix} Y(E+L_c) & Y(E+L_c-\tau) & \cdots & Y(E+L_c-(E-1)\tau) \\ Y(E+1+L_c) & Y(E+1+L_c-\tau) & \cdots & Y(E+1+L_c-(E-1)\tau) \\ \vdots & \vdots & \ddots & \vdots \\ Y(L) & Y(L-\tau) & \cdots & Y(L-(E-1)\tau) \end{bmatrix}, \tag{7}$$

$$M_X^{Lc} = \begin{bmatrix} x(E) & x(E-\tau) & \cdots & x(E-(E-1)\tau) \\ x(E+1) & x(E+1-\tau) & \cdots & x(E+1-(E-1)\tau) \\ \vdots & \vdots & \ddots & \vdots \\ x(L-L_c) & x(L-L_c-\tau) & \cdots & x(L-L_c-(E-1)\tau) \end{bmatrix}, \tag{8}$$

where Lc is the preset number of causal lag days; E is the embedding dimension (determined by the pseudo-nearest-neighbor method combined with empirical evidence, and $\mathrm{E} = 3$ in this study); τ is the time lag ($\tau = 1$ in this study); $Y(t)$ denotes the value of variable Y at time t, where $t \in [E, L]$; M is the reconstructed manifold matrix, and $M_Y \in \mathbb{R}^{(L-Lc-(E-1)\tau)\times E}$, with each row representing an embedding vector. Since the two matrices have the same dimensions, cross-mapping predictions can be conducted row-by-row.

For each valid time t, the algorithm computes the Euclidean distance d[y(i), y(j)] between any pair of phase-space points. It identifies the indices Nt of the K-nearest neighbors of y(t) on M_Y^{Lc} ($Nt = \{t_1, t_2, ..., t_{E+1}\}$), and derives the exponential weights ω_i:

$$\mu_i = \exp\left(-\frac{d[y(t),y(t_i)]}{d[y(t),y(t_1)]}\right), i = 1, \ldots, E+1\ , \tag{9}$$

$$\omega_i = \frac{\mu_i}{\sum_{j=1}^{E+1} \mu_j}\ , \tag{10}$$

where $d[y(t), y(t_1)]$ denotes the nearest-neighbor distance; μ_i is the distance weight of the i-th nearest neighbor; ω_i represents the final normalized exponential weight for the i-th neighbor; the number of neighborhood points K was set to E+1 (Agnon et al., 1999; Sugihara and May, 1990), a standard setting in the CCM method, to ensure the minimum degrees of freedom required for solving the linear equation system.

With $X(t)$ as the prediction target, the reconstructed value of $X(t)$ is calculated via the following formula:

$$\hat{X}(t)|M_Y(t+L_c) = \sum_{i=1}^{E+1} \omega_i \cdot X(t_i)\ , \tag{11}$$

Where $\hat{X}(t)$ represents the predicted value derived from the shadow manifold M_Y under a given lag Lc; $X(t_i)$ denotes the observed scalar value of variable X at time t corresponding to the i-th nearest neighbor vector.

This prediction routine runs across all valid time indices to produce the predicted series $\hat{X}$ and the original observed series X. It then calculates the Pearson correlation coefficient as the cross-mapping predictive skill ρ, and examined the convergence of ρ with respect to the library length L across varying L. The global scanning algorithm identifies the optimal causal lag Lc,opt of X relative to Y.

To test the statistical significance of the time-lagged cross-mapping results, we adopted the surrogate data method (Krishna and Tangirala, 2019), to conduct significance tests for each Lc.

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

2.2.4 Validation Protocol

To validate the effectiveness of the L-CCM method, we designed a two-stage validation approach, utilizing monthly-scale and daily-scale data respectively, to assess the improvement in both the identification of macro-level causal relationships and the quantification of optimal time lags.

First, the CCM method was applied to monthly-scale data from 2021 to obtain baseline results regarding the causal relationships between each variable X and pCO_2 at the macro scale, thereby establishing a set of candidate factors and determining the prior causal direction. These baseline results were used as a reference standard for subsequent comparisons to assess the accuracy of causal identification and the extent of performance improvement brought by the introduction of time lags.

Then, the L-CCM method was adopted to analyze the daily data, yielding the optimal causal time lag Lc,opt for each driving factor and the causal strength considering the time lag.

By comparing the L-CCM results with the CCM results, the performance improvement was evaluated from the following three aspects:

(a) Improved causal detection ability: Can the L-CCM framework capture causal interactions that standard CCM fails to identify?

(b) Enhanced causal strength: Does L-CCM produce significantly larger causal strength values compared with standard CCM?

(c) Quantification of causal time lags: Can the L-CCM method effectively quantify the time lag of the driving factors' influence on the response variable?

3 Results and Discussion

3.1 CCM Causal Identification Results

The CCM method was used to preliminarily identify the causal relationships between each potential driver X and sea surface pCO_2. The analysis was based on monthly mean data spanning 21 years (2000–2020), covering a total of 18 potential drivers.

3.1.1 Preliminary Identification of Causal Relationships

We identified the causal relationships between the sea surface pCO_2 and each X, and derived their bidirectional causal strength ρ, as visualized in Fig. 1.

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

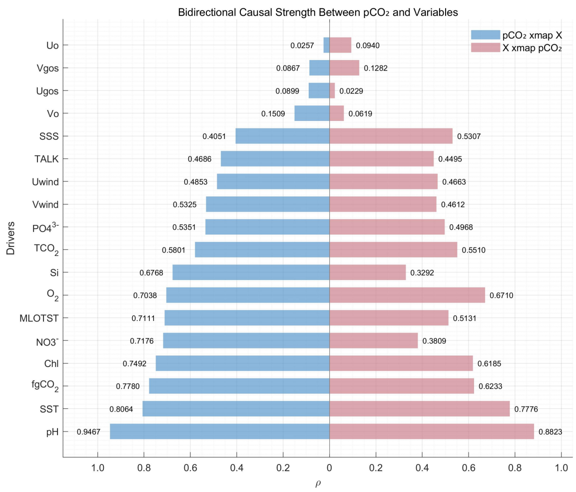

Figure 1. Causal strength statistics between pCO_2 and *X*. Causal strength statistics between sea surface pCO_2 and *X*. y xmap x represents the cross-mapping ability from y to x, which reflects the causal strength of x acting on pCO_2; x xmap y represents the cross-mapping ability from x to pCO_2, indicating the reverse causal influence of pCO_2 on x.

The results indicated that causal relationships existed between most potential drivers *X* and sea surface pCO_2. Among these, *p*H and SST exhibited the highest cross-mapping skill with pCO_2, with ρ values reaching 0.9467 (pCO_2 xmap *p*H) and 0.8064 (pCO_2 xmap SST), respectively, indicating a tight coupling relationship between these factors and pCO_2. In addition, the cross-mapping coefficients for $fgCO_2$ (0.7780/0.6233), Chl (0.7492/0.6185), NO_3^- (0.7176/0.3809), MLOTST (0.7111/0.5131), and O_2 (0.7038/0.6710) were overall high, indicating significant causal relationships between these factors and pCO_2. This confirmed that the dominant causal direction was *X*→Y, i.e., these potential drivers *X* influenced pCO_2.

In comparison, *U*o (0.0257/0.094), *V*o (0.1509/0.0619), *U*gos (0.0899/0.0229), *V*gos (0.0867/0.1282), and other factors exhibited lower cross-mapping coefficients, some even approaching zero, suggesting that these factors exhibited weak causal relationships with pCO_2 or virtually no causal linkage.

Overall, physical factors (SST, MLOTST, wind speed, etc.), chemical factors (*p*H, TCO_2, NO_3^-, etc.), and biological factors (Chl) all exerted driving effects of varying degrees on sea surface pCO_2, which was consistent with the fundamental understanding of physical-chemical-biological multi-process coupling in the ocean carbon cycle (Lekshmi et al., 2023).

3.1.2 Convergence Test and Causal Direction Determination

Figure 2 shows the cross-mapping skill ρ between each potential driver *X* and pCO_2 as a function of sequence length L. Overall, most variable pairs showed a typical convergence trend with increasing L: ρ(L) rose gradually and eventually plateaued, satisfying the basic convergence prerequisites of CCM. Taking SST and pCO_2 as an example, the result showed that as L increased to the maximum available length, ρ(pCO_2 xmap SST) increased from zero and stabilized at approximately 0.80, while ρ(SST xmap pCO_2) rose steadily and plateaued near 0.77, with an obvious convergent pattern. Moreover, all

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

significance test p-values were lower than 0.05, indicating that the causal relationship between SST and pCO_2 was statistically reliable and satisfied the convergence requirements of the CCM method. A bidirectional coupling relationship can therefore be inferred between SST and pCO_2, which was consistent with the complex interactive features of the ocean carbon cycle system. In this coupling, SST exerted a strong and dominant causal control on pCO_2, whereas pCO_2 also presented an evident inverse regulatory effect; this reverse signal may partly reflect data coupling or spurious feedback induced by strong primary causal interactions.

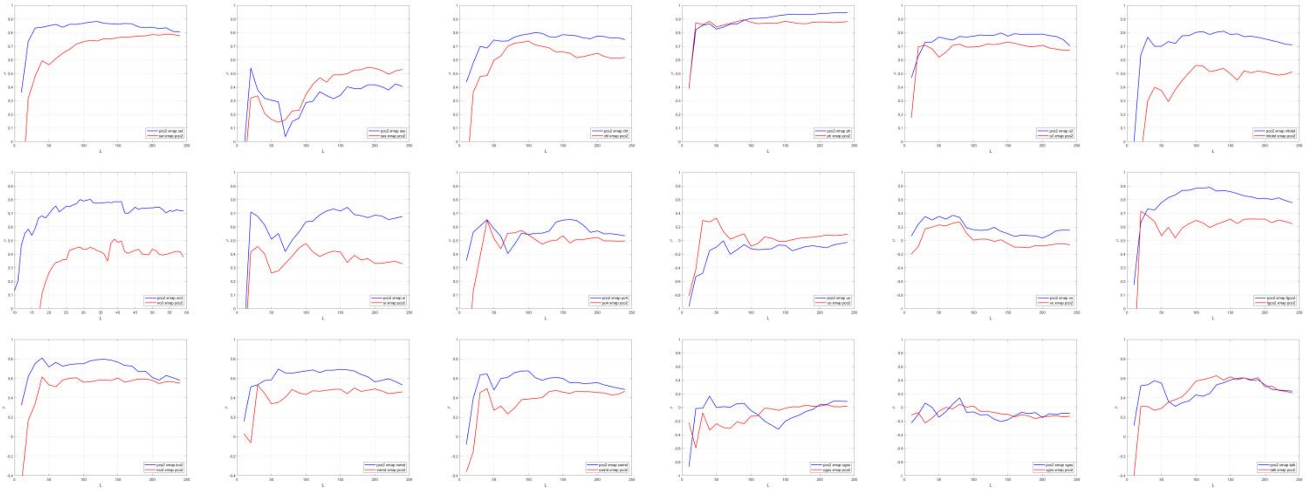

Figure 2. Cross-mapping convergence curves. The blue curve in the figure represents the causal strength of potential driver X on pCO_2; the red curve represents the reverse causal strength of pCO_2 on potential driver X; the horizontal axis represents the library length L; the vertical axis represents the cross-mapping skill ρ.

The convergence test results for other major driving factors (e.g., pH, Chl) were similar. As the time series length L increased, the cross-mapping prediction accuracy generally rose and gradually stabilized. For some factors with relatively low ρ values (e.g., *U*wind, *V*wind), although their convergence curves displayed a slight upward trend, convergence values remained low with obvious fluctuations, further confirming their relatively weak causal relationships with pCO_2.

For several additional factors (e.g., *V*o and *U*o), the convergence curves showed nearly no upward trend, with ρ values fluctuating around zero. These factors failed to satisfy the CCM convergence criteria and were thus determined to have no causal association with pCO_2.

In most cases, the ρ values for the reverse mapping ($pCO_2 \rightarrow X$) were lower than those for the forward mapping, with poor convergence; this thus supported the directional judgment of $X \rightarrow Y$. With the exception of SSS, the reverse ρ ($pCO_2 \rightarrow X$) was slightly higher than those of the forward mapping and passed the significance test, indicating that pCO_2 played a dominant role in the relationship between these two variables. This finding differed from the classical understanding within the ocean-atmosphere carbon exchange framework (Weiss et al., 1982) and thus warranted further analysis.

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

3.1.3 Limitations of Existing Methods

Based on the above results, most factors *X* showed strong cross-mapping ability with pCO_2 (ρ generally above 0.6) and good convergence performance, indicating that these factors exerted significant causal effects on pCO_2 with clear directional characteristics. Notably, several factors *X* (e.g., SSS and TALK) were theoretically closely coupled with pCO_2 in physical mechanisms, yet they presented relatively low ρ values in the monthly-scale CCM analysis. Taking SSS as an example, the ρ value of pCO_2 xmap SSS was only 0.40, considerably lower than that of SST (0.80) and other key factors, and it did not dominate the corresponding causal relationship. This phenomenon warranted further investigation: whether such factors intrinsically impose weaker controls on pCO_2, or whether the monthly temporal resolution led to systematic underestimation and biased causal interpretation.

From the perspective of ocean carbon sink mechanisms, the latter possibility cannot be neglected. In ocean carbon cycling processes, the responses induced by driving factors typically present time lags of several days to weeks, rather than instantaneous interactions. Most time-series datasets adopted in standard CCM causal analyses are aggregated at monthly, annual or longer timescales (Wang et al., 2018), which may mix and obscure lag-dependent signals. Averaging high-frequency variations over coarse temporal resolutions inevitably smooths out lag information, thereby lowering estimated ρ values and weakening the overall accuracy of cross-mapping inference. This reflects an inherent limitation of standard CCM when applied to time-lagged biogeochemical interactions, and further motivates the incorporation of explicit causal lag parameters to refine the classical CCM framework.

3.2 L-CCM Causal Identification Results

3.2.1 Identification and Characteristic Analysis of $L_{c,\mathrm{opt}}$

Based on the analysis of CCM limitations in Sect. 3.1.3, we employed the L-CCM method to conduct causal analysis, determine the optimal causal time lag and assess its improvement in identification accuracy. Relevant results are shown in Fig. 3 and Fig. 4.

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

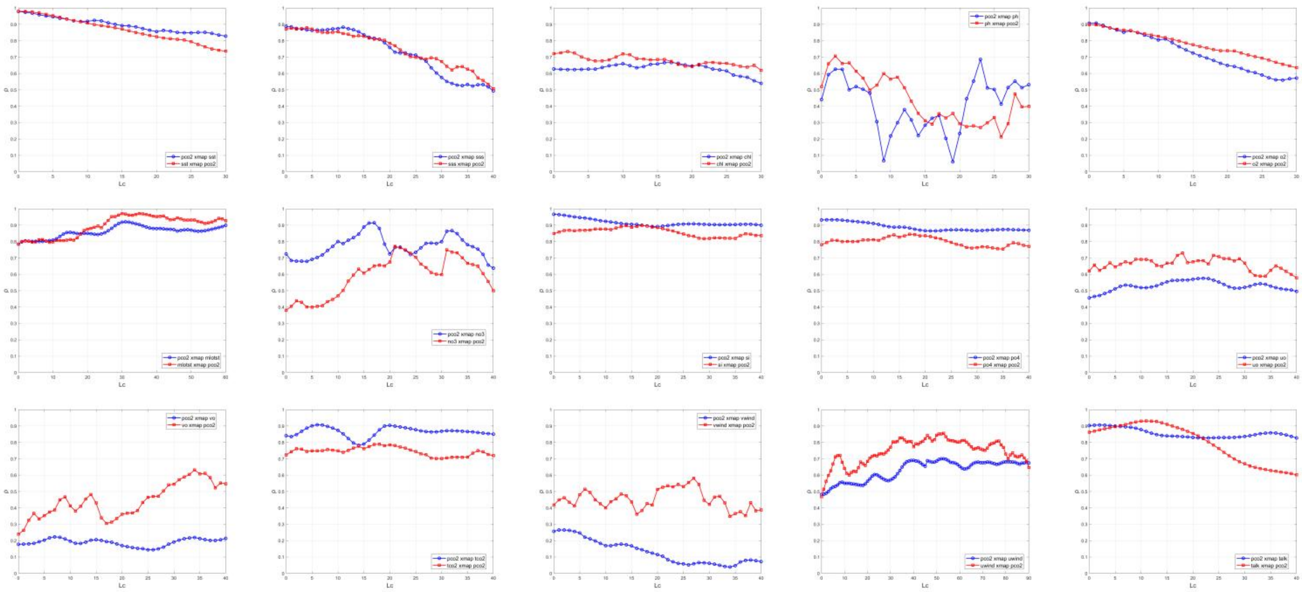

Figure 3. Identification results of optimal time lags. The ρ values denote cross-mapping skill, indicating the causal strength.

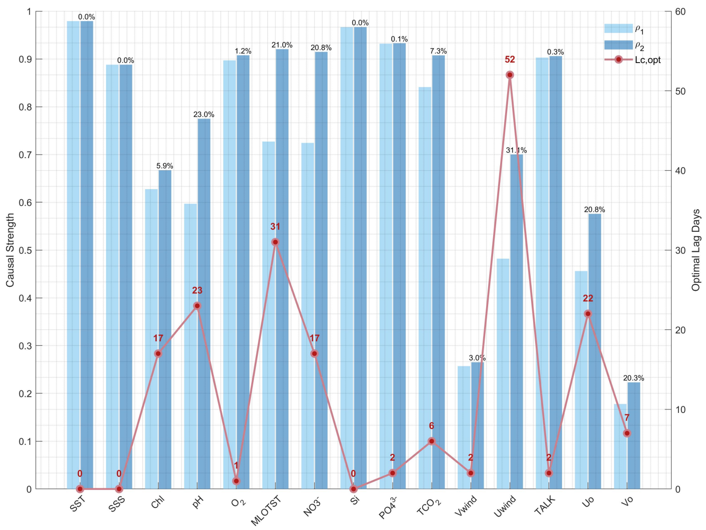

Figure 4. Improvement of Causal Identification by L-CCM. *X* represents the potential drivers of $p\mathrm{CO_2}$. ρ_1 refers to causal strength without considering time lag when $Lc = 0$; ρ_2 denotes causal strength calculated via the L-CCM method. Lc,opt is the optimal causal time lag. The labeled percentages indicate the accuracy improvement range of the L-CCM approach.

Based on the results in Fig. 3, the impact of different types of potential drivers on $p\mathrm{CO_2}$ showed significant temporal variations. According to their time-lag characteristics, these drivers can be divided into three categories:

(a) Short-lag drivers (Lc,opt ≤ 7 days):

These include SST (0 days), SSS (0 days), Si (0 days), $\mathrm{O_2}$ (1 day), $\mathrm{PO_4^{3-}}$ (2 days), *V*wind (2 days), TALK (2 days), and $T\mathrm{CO_2}$ (6 days).

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

The significant causal effects of these factors peak within only a few days, reflecting that their impacts on pCO_2 are dominated by rapid physical mixing processes (e.g., the control of temperature on pCO_2 solubility and the instantaneous effect of salinity on seawater stratification) and rapid chemical responses. Their modulating effects on pCO_2 show nearly no time lag or only a short delay of several days. The short-lag characteristics of these drivers are consistent with previous findings that physical forcings (e.g., wind and heat flux) and rapid biogeochemical processes dominate surface seawater carbon variations at the daily timescale (Byju and Prasanna Kumar, 2011; Lu et al., 2012; Merlivat et al., n.d.; Thomas et al., 2012).

(b) Medium-to long-lag drivers (Lc,opt > 7 days):

These include Chl (17 days), NO_3^-(17 days), Uo (22 days), pH (23 days), and MLOTST (31 days). The influences of these drivers present time lags of several weeks, which are primarily attributed to biological responses in biogeochemical processes (e.g., the biological pump effect associated with phytoplankton growth, physical accumulation, and cross-scale transport) (DeVries, 2022; Stange et al., 2017; Wang et al., 2025), reflecting the slow-response characteristics of ocean dynamic processes.

(c) Factors with weak causal relationships:

Although the ρ values for Uo and Vo showed a slight increase in the daily-scale time-lag scan, they remained low (0.45 and 0.17, respectively). Overall, no significant causal signals were identified, which was consistent with the weak causal results obtained from CCM identification.

3.2.2 Improvement of Causal Identification by L-CCM

(a) Significant Improvement in Causal Strength

We used the L-CCM method to identify causal relationships between pCO_2 and its potential drivers. Compared with the results obtained without considering time lags, this method achieved a distinct improvement in performance. As shown in the results of Fig. 4, the causal strength of the vast majority of potential drivers X was improved to varying degrees after introducing the optimal time lag, which fully validated the optimization effect of this method on causal identification.

For example, the ρ value of pH was 0.5969 at $Lc = 0$; after incorporating the optimal lag of 23 days, ρ increased to 0.7748, with an improvement of approximately 23 %. For MLOTST, ρ was 0.7271 at $Lc = 0$; after incorporating the optimal lag of 31 days, ρ rose to 0.9205, representing a 21 % improvement. For NO_3^-, ρ was 0.7243 at $Lc = 0$; after incorporating the optimal lag of 17 days, ρ increased to 0.9148, showing a 21 % improvement. For Chl, ρ was 0.6275 at $Lc = 0$; after incorporating the optimal lag of 17 days, ρ increased to 0.6671, with a relatively small but still notable improvement of 6 %. It is evident that the causal strength of most potential drivers increased to varying degrees, with an average growth of 10.32 %.

Several drivers (e.g., SST, SSS, and Si) already yielded relatively high ρ values (>0.8) at $Lc = 0$, with their optimal time lags fixed at 0 days. Accordingly, their causal strength ρ remained unchanged under the L-CCM framework, indicating that these factors exert instantaneous or near-instantaneous impacts on pCO_2.

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

(b) Comparison with Monthly-Scale Results

Overall, even without time lags ($Lc = 0$), daily-scale results outperformed monthly-scale outcomes. For instance, the ρ of SSS increased from 0.4051 (monthly, non-dominant causality) to 0.8880 (daily, dominant causality), a 54 % rise. Si rose from 0.6768 to 0.9666 (30 %), and TALK from 0.4686 to 0.9028 (48 %). All other drivers also improved. This supported the inference in Sect. 3.1.3. Low-resolution monthly data average time-lag signals, causing systematic underestimation and misjudgment of causal strength. Higher-resolution daily data captures subtle temporal variations, effectively improving the causal identification accuracy of CCM.

A comparison between the two sets of experimental results revealed consistent overall tendencies of the two methods. All causal relationships identified at the monthly-scale were maintained with daily-scale data and optimal time lags. Drivers with underestimated causality in monthly results, such as SSS and TALK, exhibited enhanced causal strength after identification via the L-CCM, which further confirmed their causal effects and reduced misjudgment risks. The L-CCM method can accurately quantify time-lag characteristics, providing critical support for in-depth analysis of causal mechanisms in ocean carbon cycling.

3.2.3 Special Case Analysis and Discussion of Method Limitations

Chl, pH and Vwind showed lower ρ values at the daily-scale ($Lc = 0$) than at the monthly-scale, which may be attributed to insufficient data volume. For instance, the pH time series lacked continuous records; only 80 valid continuous samples were retained after preprocessing. The limited sample size might lead to unstable results and incomplete convergence. Nevertheless, the ρ values of these variables increased after adopting the optimal time lag, demonstrating that the introduction of Lc exerted a positive effect on improving causal identification.

The optimal time lag of Uwind was 52 days, exceeding the preset scanning range of 40 days. There are three possibilities:

(a) the influence of wind fields on pCO_2 involves a long cascading delay (e.g., wind → mixed layer variation → nutrient supply → biological response → pCO_2) (Jones et al., 2014), though a 52-day lag is unreasonably long;

(b) the length of daily-scale data is insufficient to support stable identification of such long time lags;

(c) Uwind is dominated by seasonal effects, which mismatch the research objective of daily-scale analysis.

Identical analytical methods applied to the same dataset produce no significant discrepancies in repeated analyses, except for recording errors.Variables with unstable identification outcomes were thus excluded from the main conclusions. This revealed inherent limitations of the CCM method, which struggles to deliver definite and stable causal judgements for partial results.

4 Conclusions

This study first adopted CCM to identify potential drivers of sea surface pCO_2 in the subtropical Northwest Pacific. Multiple significant drivers were recognized, including pH, SST, $fgCO_2$, Chl, NO_3^-, MLOTST, O_2, Si, TCO_2 and PO_4^{3-}. These factors constituted a sophisticated nonlinear regulatory network with obvious differences in bidirectional causal coupling. Driving factors generally exhibited stronger causal impacts on pCO_2 compared with the reverse feedback from pCO_2 itself. However, the standard CCM failed to account for time-lag effects, resulting in neglected and underestimated contributions of drivers such as SSS and TALK. To overcome this deficiency, this work incorporated the causal time-lag parameter *Lc* and developed the L-CCM model accordingly.

Results obtained from the L-CCM analysis indicated that all drivers exhibit distinct causal time-lag characteristics. Sea surface pCO_2 responds rapidly to short time-lag drivers including SST, SSS, Si and O_2, reflecting rapid physicochemical regulation processes such as thermal variation and seawater carbonate balance in the study area. In contrast, delayed responses associated with medium and long time-lag drivers like Chl, NO_3^- and *p*H are mainly controlled by slow oceanic processes, including phytoplankton succession and the evolution of the upper mixed layer. The differences in time lag further demonstrate that the evolution of sea surface pCO_2 results from the coupling of multi-scale physical, chemical and biological processes. Apart from identifying the optimal causal time lag of each driver, this method also captured drivers underestimated by the standard CCM approach, effectively improving the accuracy of causal identification. The results suggested that the prominent causal drivers of sea surface pCO_2 in the study region cover SST, SSS, Chl, *p*H, O_2, MLOTST, NO_3^-, Si, PO_4^{3-}, TCO_2 and TALK, whereas *U*o, *V*o, *U*gos, *V*gos and *V*wind present weak causal relationships. Among them, SST, Si, PO_4^{3-}, O_2, TCO_2 and TALK serve as core dominant drivers of pCO_2 variation. With high causal strength and stability, they exert the most significant nonlinear regulation effects on pCO_2. MLOTST, NO_3^-, SSS, Chl and *p*H are classified as important secondary drivers.

Compared with the standard CCM method, the L-CCM model greatly optimized the detection capability of asynchronous causal relationships, and effectively compensated for deficiencies such as underestimated causal strength and ambiguous causal directions caused by neglected time lag effects. This method adapts to the multi-scale and asynchronous evolutionary features of ocean carbon cycle. It can accurately distinguish time-lag differences among various environmental drivers, and the identification results achieve higher precision and better interpretability of physical mechanisms. The established causal identification framework is not only suitable for analyzing the ocean carbon cycle mechanisms, but also possesses favorable generality. It is expected to offer methodological references for causal inference studies in Earth system science and other complex dynamic systems.

Author contributions

MQ, XC, and SJ collected the datasets of surface ocean pCO_2 and its potential driving factors. YW and ZL conducted dataset integration and preprocessing. HW and ML designed the Lag-Convergent Cross Mapping (L-CCM) model and implemented the identification of driving factors for surface ocean pCO_2. ML, JL, and XL synthesized and analyzed the results and further refined the algorithm. ML prepared the manuscript with contributions from all co-authors.

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

Competing interests

The corresponding author declares that neither the authors nor their affiliated institutions have any competing interests.

Acknowledgements

We acknowledge the Surface Ocean CO_2 Atlas (SOCAT) consortium for global quality-controlled fCO_2 in situ measurements, NASA Ocean Color for Level-3 satellite data, and CMEMS for Level-4 reanalysis gridded products. Special thanks go to the Kuroshio Extension Observatory (KEO) for long-term daily mooring observations, which enabled our time-lag causality analysis of sea surface pCO_2. We thank the editor and anonymous reviewers for their constructive suggestions to improve this work.

Financial support

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by Key Laboratory of Land Satellite Remote Sensing Application, Ministry of Natural Resources of the People' s Republic of China, grant numbers G202211.

References

Agnon, Y., Golan, A., and Shearer, M.: Nonparametric, nonlinear, short-term forecasting: theory and evidence for nonlinearities in the commodity markets, Econom. Lett., 65, 293–299, https://doi.org/10.1016/S0165-1765(99)00156-1, 1999.

Kuroshio extension observatory | ocean climate stations: https://www.pmel.noaa.gov/ocs/KEO, last access: 7 April 2026.

NASA ocean color: https://oceancolor.gsfc.nasa.gov/, last access: 7 April 2026.

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S. R., Balestrini, C. F., Barbero, L., Bates, N. R., Bianchi, A. A., Bonou, F., Boutin, J., Bozec, Y., Burger, E. F., Cai, W.-J., Castle, R. D., Chen, L., Chierici, M., Currie, K., Evans, W., Featherstone, C., Feely, R. A., Fransson, A., Goyet, C., Greenwood, N., Gregor, L., Hankin, S., Hardman-Mountford, N. J., Harlay, J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Huss, B., Ibánhez, J. S. P., Johannessen, T., Keeling, R., Kitidis, V., Körtzinger, A., Kozyr, A., Krasakopoulou, E., Kuwata, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lo Monaco, C., Manke, A., Mathis, J. T., Merlivat, L., Millero, F. J., Monteiro, P. M. S., Munro, D. R., Murata, A., Newberger, T., Omar, A. M., Ono, T., Paterson, K., Pearce, D., Pierrot, D., Robbins, L. L., Saito, S., Salisbury, J., Schlitzer, R., Schneider, B., Schweitzer, R., Sieger, R., Skjelvan, I., Sullivan, K. F., Sutherland, S. C., Sutton, A. J., Tadokoro, K., Telszewski, M., Tuma, M., van Heuven, S. M. A. C., Vandemark, D., Ward, B., Watson, A. J., and Xu, S.: A multi-decade record of high-quality fCO2 data in version 3 of the surface ocean CO2 atlas (SOCAT), Earth Syst. Sci. Data, 8, 383–413, https://doi.org/10.5194/essd-8-383-2016, 2016.

Byju, P. and Prasanna Kumar, S.: Physical and biological response of the arabian sea to tropical cyclone phyan and its implications, Mar. Environ. Res., 71, 325–330, https://doi.org/10.1016/j.marenvres.2011.02.008, 2011.

Chen, S., Hu, C., Barnes, B. B., Wanninkhof, R., Cai, W.-J., Barbero, L., and Pierrot, D.: A machine learning approach to estimate surface ocean pCO2 from satellite measurements, Remote Sens. Environ., 228, 203–226, https://doi.org/10.1016/j.rse.2019.04.019, 2019.

Chen, T., Cao, Z., Stewart, J., and Wan, X.: Preface to the special issue: western pacific biogeochemical cycles across the air-sea and sediment-water interfaces and their response to global changes, Global Planet. Change, 254, 105013, https://doi.org/10.1016/j.gloplacha.2025.105013, 2025.

DeGrandpre, M. D., Wanninkhof, R., McGillis, W. R., and Strutton, P. G.: A Lagrangian study of surface p CO2 dynamics in the eastern equatorial Pacific Ocean, J. Geophys. Res., 109, 2003JC002089, https://doi.org/10.1029/2003JC002089, 2004.

DeVries, T.: The ocean carbon cycle, Annu. Rev. Environ. Resour., 47, 317–341, https://doi.org/10.1146/annurev-environ-120920-111307, 2022.

Jones, D. C., Ito, T., Takano, Y., and Hsu, W.: Spatial and seasonal variability of the air-sea equilibration timescale of carbon dioxide, Global Biogeochemical Cycles, 28, 1163–1178, https://doi.org/10.1002/2014GB004813, 2014.

Katavouta, A. and Williams, R. G.: Ocean carbon cycle feedbacks in CMIP6 models: contributions from different basins, Biogeosciences, 18, 3189–3218, https://doi.org/10.5194/bg-18-3189-2021, 2021.

Krishna, P. V. and Tangirala, A. K.: Inferring direct causality from noisy data using convergent cross mapping, in: 2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 1523–1528, https://doi.org/10.23919/SICE.2019.8859922, 2019.

Lekshmi, K., Bharti, R., and Mahanta, C.: Significance of physico-chemical and biological parameters on spatio-temporal PCO2 distribution in the northeastern Indian Ocean, J. Indian Soc. Remote Sens., 51, 2511–2519, https://doi.org/10.1007/s12524-023-01780-3, 2023.

Liu, S., Liang, J., Jiang, Z., Li, J., Wu, Y., Fang, Y., Ren, Y., Zhang, X., Huang, X., and Macreadie, P. I.: Temporal and spatial variations of air-sea CO2 fluxes and their key influence factors in seagrass meadows of hainan island, south China sea, Sci. Total Environ., 910, 168684, https://doi.org/10.1016/j.scitotenv.2023.168684, 2024.

López-Puertas, A., Wurl, O., Frka, S., and Ribas-Ribas, M.: Diel variability affects the inorganic carbon system in the sea-surface microlayer and influences air-sea CO2 flux estimates, Ocean Sci., 21, 3471–3485, https://doi.org/10.5194/os-21-3471-2025, 2025.

Lu, Z., Gan, J., and Dai, M.: Modeling seasonal and diurnal pCO2 variations in the northern south China sea, J. Mar. Syst., 92, 30–41, https://doi.org/10.1016/j.jmarsys.2011.10.003, 2012.

https://doi.org/10.5194/egusphere-2026-3875
Preprint. Discussion started: 2 September 2026

Merlivat, L., Davila, M. G., Caniaux, G., Boutin, J., and Reverdin, G.: Mesoscale and diel to monthly variability of CO2 and carbon fluxes at the ocean surface in the northeastern atlantic, https://doi.org/10.1029/2007JC004657, n.d.

Randall, C. J. and van Woesik, R.: Some coral diseases track climate oscillations in the caribbean, Sci. Rep., 7, 5719, https://doi.org/10.1038/s41598-017-05763-6, 2017.

Stange, P., Bach, L. T., Le Moigne, F. a. C., Taucher, J., Boxhammer, T., and Riebesell, U.: Quantifying the time lag between organic matter production and export in the surface ocean: implications for estimates of export efficiency, Geophys. Res. Lett., 44, 268–276, https://doi.org/10.1002/2016GL070875, 2017.

Sugihara, G. and May, R. M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, Nature, 344, 734–741, https://doi.org/10.1038/344734a0, 1990.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., and Munch, S.: Detecting causality in complex ecosystems, Science, 338, 496–500, https://doi.org/10.1126/science.1227079, 2012.

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J., and Nojiri, Y.: Global sea–air CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects, Deep Sea Res. Part II, 49, 1601–1622, https://doi.org/10.1016/S0967-0645(02)00003-6, 2002.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R., and de Baar, H. J. W.: Climatological mean and decadal change in surface ocean pCO2, and net sea–air CO2 flux over the global oceans, Deep Sea Res. Part II, 56, 554–577, https://doi.org/10.1016/j.dsr2.2008.12.009, 2009.

Thomas, H., Craig, S. E., Greenan, B. J. W., Burt, W., Herndl, G. J., Higginson, S., Salt, L., Shadwick, E. H., and Urrego-Blanco, J.: Direct observations of diel biological CO2 fixation on the scotian shelf, northwestern Atlantic Ocean, Biogeosciences, 9, 2301–2309, https://doi.org/10.5194/bg-9-2301-2012, 2012.

Tjiputra, J. F., Couespel, D., and Sanders, R.: Marine ecosystem role in setting up preindustrial and future climate, Nat. Commun., 16, 2206, https://doi.org/10.1038/s41467-025-57371-y, 2025.

Tozawa, M., Nomura, D., Matsuura, M., Hatta, M., Fujiwara, A., Yasunaka, S., and Murata, A.: Quantitative assessment of factors contributing to variations in sea surface pCO2 in the pacific sector of the Arctic Ocean, J. Geophys. Res.: Oceans, 129, e2024JC021012, https://doi.org/10.1029/2024JC021012, 2024.

Wang, G., Dai, M., Shen, S. S. P., Bai, Y., and Xu, Y.: Quantifying uncertainty sources in the gridded data of sea surface CO2 partial pressure, J. Geophys. Res.: Oceans, 119, 5181–5189, https://doi.org/10.1002/2013JC009577, 2014.

Wang, R., Li, X., Song, J., Wang, Z., Zhong, G., Yuan, H., and Duan, L.: Surface seawater chlorophyll-a variability in the south China sea: influence of pCO2 and co-varying environmental factors, Environmental Research, 279, 121808, https://doi.org/10.1016/j.envres.2025.121808, 2025.

Wang, Y., Yang, J., Chen, Y., De Maeyer, P., Li, Z., and Duan, W.: Detecting the causal effect of soil moisture on precipitation using convergent cross mapping, Sci. Rep., 8, 12171, https://doi.org/10.1038/s41598-018-30669-2, 2018.

Weiss, R. F., Jahnke, R. A., and Keeling, C. D.: Seasonal effects of temperature and salinity on the partial pressure of CO2 in seawater, Nature, 300, 511–513, https://doi.org/10.1038/300511a0, 1982.

Zhong, G., Li, X., Song, J., Qu, B., Wang, F., Wang, Y., Zhang, B., Sun, X., Zhang, W., Wang, Z., Ma, J., Yuan, H., and Duan, L.: Reconstruction of global surface ocean pCO2 using region-specific predictors based on a stepwise FFNN regression algorithm, Biogeosciences, 19, 845–859, https://doi.org/10.5194/bg-19-845-2022, 2022.

Zhou, C., Li, H., Hu, Y., Zhang, B., Ren, P., Kan, Z., Jia, X., Mi, J., and Guo, X.: Causal effects of key air pollutants and meteorology on ischemic stroke onset: a convergent cross-mapping approach, Ecotoxicol. Environ. Saf., 291, 117861, https://doi.org/10.1016/j.ecoenv.2025.117861, 2025.

Zou, Z., Cheng, C., and Shen, S.: The complex nonlinear coupling causal patterns between PM2.5 and meteorological factors in Tibetan Plateau: a case study in Xining, IEEE Access, 9, 150373–150382, https://doi.org/10.1109/ACCESS.2021.3123455, 2021.