

Supplementary Information

Monitoring, modeling, and forecasting long-term changes in coastal seawater quality due to climate change

Xianghong Guan^{a,#}, Hua Huang^{a,#}, Xiong Ke^a, Xiaoqian Cheng^a, Heng Zhang^a, Acong Chen^a, Guanglei Qiu^a, Haizhen Wu^b, Chaohai Wei^{a,c,*}

^a School of Environment and Energy, South China University of Technology, Guangzhou 510006, P. R. China

^b School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, P. R. China

^c School of Environment and Energy, Guangdong Provincial Key Laboratory of Solid Wastes Pollution Control and Recycling, South China University of Technology, Guangzhou 510006, P. R. China

[#]The authors contributed equally to this work.

^{*}Corresponding author email address:

Chaohai Wei, cechwei@scut.edu.cn

Part A The collection and collation of measured data

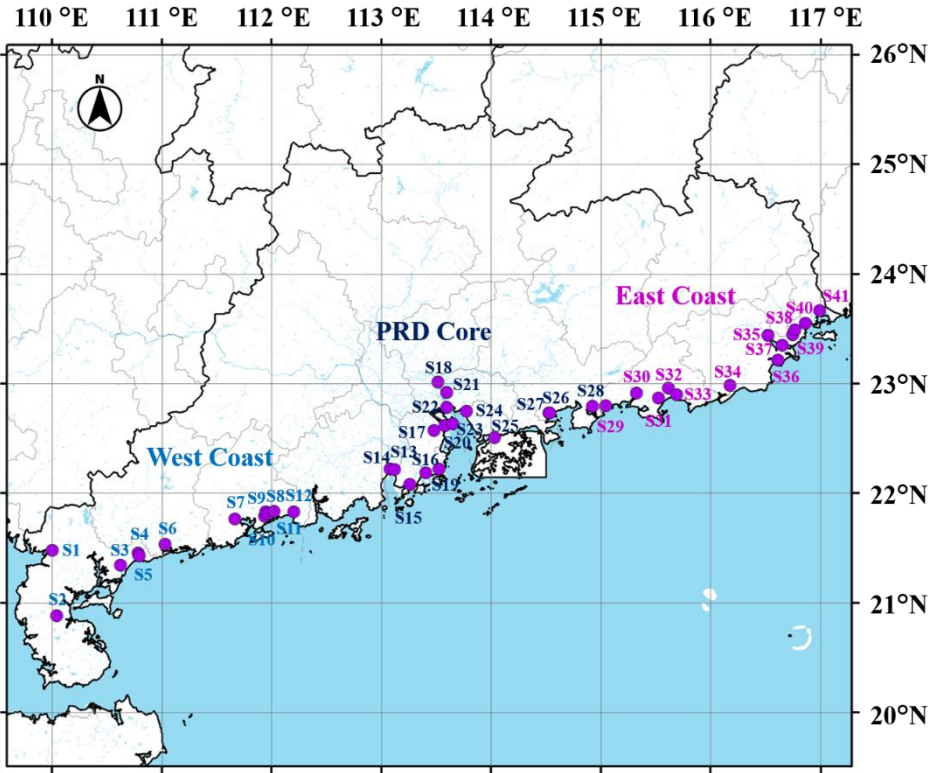
SA.1 Methods of sample collection and analysis

The positions of 41 sampling points along the Guangdong coast are illustrated in Supplementary **Fig. 1**, with their coordinates listed in Supplementary **Table 1**. Surface water was collected at these 41 sampling points using borosilicate glass bottles from January 2015 to June 2022. On-site measurements of pH, dissolved oxygen (DO), and electrical conductivity were conducted. The pH values (± 0.02) were determined using the glass electrode method (GB 11913-89, HANNA H19125), while electrical conductivity was measured using a conductivity meter (HQ1140), and DO was measured using the electrochemical probe method (GB 11913-89, HACH HQ 30).

Prior to analysis, water samples were filtered through pre-treated glass fiber membranes (pore size = 45.300 μm). Chemical oxygen demand (COD) was analyzed using the Potassium chromate method (GB 11914-89, WB 800), while biochemical oxygen demand (BOD) was determined using the dilution and seeding method (GB7488-87). Total phosphorus (TP) was analyzed using the Ammonium molybdate spectrophotometric method (GB 11893-89, UNICO 2100), and total nitrogen (TN) was measured using the alkaline potassium persulfate digestion ultraviolet spectrophotometric method (GB11894-89, UNICO 2100). Ammonia nitrogen ($\text{NH}_3\text{-N}$) was determined using Nessler's Reagent spectrophotometry (GB 7479-87/UNICO 2100), while nitrate nitrogen (NO_3^-) and sulfate (SO_4^{2-}) were measured using ion chromatography (HJ/T84-2001, DX-120). Sulfide (S^{2-}) was analyzed using methyl blue spectrophotometry (GB/T 16489-1996, UNICO 2100). The concentrations of Cu, Zn, Cd, Pb, Cr(VI), Hg were determined using inductively coupled plasma mass spectrometry (ICP-MS, NexION 300x, PerkinElmer, USA), while Fe and Mn were measured using flame atomic absorption spectrophotometry (GB 11911-89).

Reagents of analytical grade or higher were used for the analysis purposes. Blank and duplicate samples were employed for quality assurance and quality control. The detection limit for Cd was $0.06 \mu\text{g L}^{-1}$, for Cr was $0.7 \mu\text{g L}^{-1}$, for Cu was $0.8 \mu\text{g L}^{-1}$, for Hg was $0.05 \mu\text{g L}^{-1}$, for Mn was 0.01 mg L^{-1} , for Fe was 0.03 mg L^{-1} , for Pb was

0.09 $\mu\text{g L}^{-1}$, and for Zn was 0.5 $\mu\text{g L}^{-1}$. The accuracy of metal determination was verified using certified reference materials (GSB04-1767-2004). Acceptance criteria included a relative standard deviation within 10% for multiple measurements and reanalysis of other samples. The variance of the standard curve in this study was greater than 0.9999. The recovery rate of dissolved heavy metals was >90%.



Supplementary Fig. 1 Geographical location of the study area and sampling sites. The shape of the water area is from the China Water Cover Map (2020/10 m)¹, and the administrative division is from ChinaAdminDivisonSHP².

Supplementary Table 1. Coordinate and region division of Sampling site

Regionalism	Sampling site	Longitude	Latitude
Coastal Economic Belt West Wing Cities	S1	110.0015	21.4831
	S2	110.0451	20.885
	S3	110.6226	21.3465
	S4	110.7848	21.45403
	S5	110.7946	21.4314
	S6	111.0329	21.5381
	S7	111.6661	21.7647
	S8	111.9399	21.7929
	S9	111.9506	21.8305
	S10	111.9695	21.81589

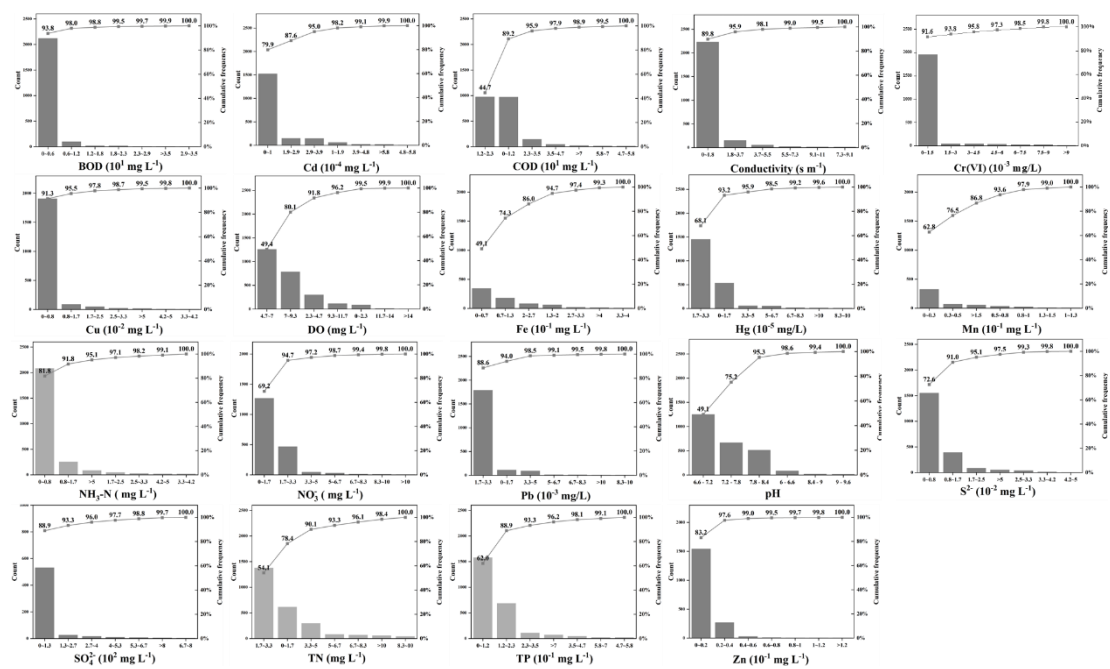
Regionalism	Sampling site	Longitude	Latitude
Coastal Economic Belt West Wing Cities	S11	112.0214	21.8353
	S12	112.2019	21.8311
	S13	113.0833	22.2248
Core Cities in the Pearl River Delta	S14	113.1235	22.2203
	S15	113.2646	22.0851
	S16	113.4092	22.1865
	S17	113.4836	22.5731
	S18	113.5157	23.0144
	S19	113.5259	22.2241
	S20	113.5816	22.6201
	S21	113.5993	22.9216
	S22	113.5997	22.7873
	S23	113.6523	22.6364
	S24	113.779	22.7518
	S25	114.0357	22.5077
	S26	114.5306	22.7362
	S27	114.5353	22.73313
	S28	114.9267	22.796
Coastal Economic Belt East Wing Cities	S29	115.0493	22.7978
	S30	115.3292	22.9156
	S31	115.5267	22.8715
	S32	115.6168	22.9585
	S33	115.6922	22.9013
	S34	116.1829	22.9834
	S35	116.5267	23.4406
	S36	116.6154	23.2187
	S37	116.6581	23.3537
	S38	116.7544	23.4478
	S39	116.7713	23.4913
	S40	116.8681	23.5537
	S41	116.9996	23.6676

Note: Urban agglomeration is divided according to the spatial distribution of the cities where the sampling sites are located and the development of economy and industry. Coastal Economic Belt West Wing Cities include Zhanjiang, Maoming, and Yangjiang, Core Cities in the Pearl River Delta include Guangzhou, Shenzhen, Zhuhai, Huizhou, Dongguan and Zhongshan, Coastal Economic Belt East Wing Cities include Shantou, Shanwei, Chaozhou and Jieyang.

SA.2 The number of effective samples for each index and Pareto diagram

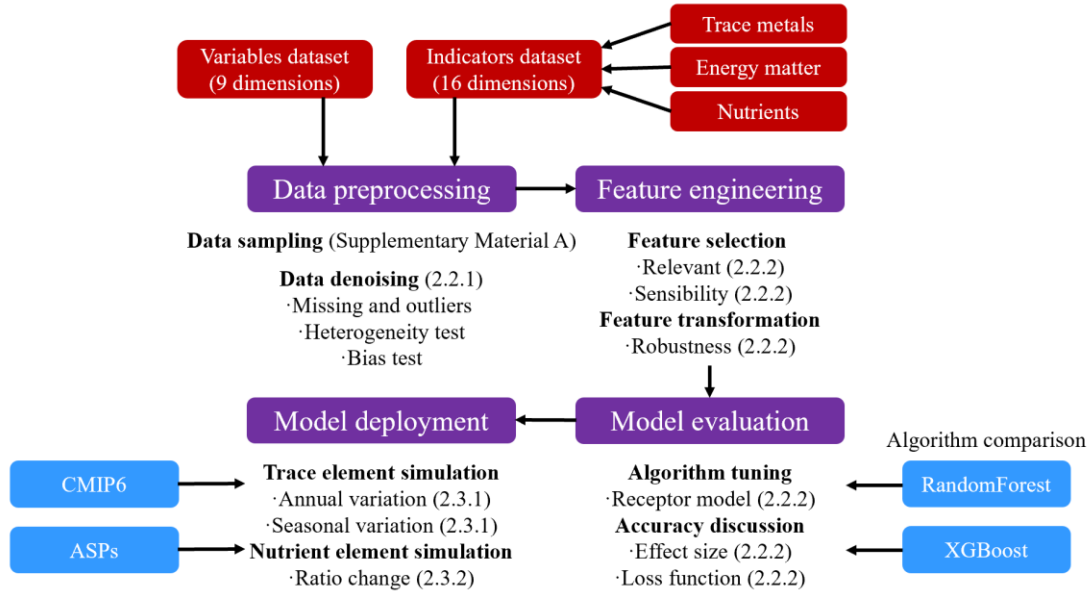
Supplementary Table 2. The number of effective samples for each index

Index	The number of effective samples
Conductivity	2481
DO	2548
pH	2549
Cd	1907
Cr(VI)	2130
Cu	2089
Fe	701
Hg	2134
Mn	514
Pb	2018
Zn	1853
BOD ₅	2260
COD	2179
NH ₃ -N	2549
NO ₃ ⁻	1831
TN	2547
TP	2548
S ²⁻	2135
SO ₄ ²⁻	597



Supplementary Fig. 2. Pareto diagram of each index.

Part B The establishment of receptor model



Supplementary Fig. 3. Model establishment and verification process.

SB1. Review of the datasets

Data below the detection limit were adjusted to a concentration that was half of the detection limit, and values outside the range of $(\bar{x} - 3\delta, \bar{x} + 3\delta)$ were treated as outliers and excluded after passing the Kolmogorov–Smirnov test. To ensure time series integrity, the number of missing values resulting from undetected and excluded data needed to be below 10%, and no interpolation of missing values was performed due to unclear spatiotemporal variation mechanisms.

Although we used temperature as discriminating tags for the datasets, further confirmation was needed to determine (1) whether the effect mechanisms of the experimental and control groups were significantly different and (2) if the difference resulted from global or individual bias data. To address these questions, we used heterogeneity and bias tests in the meta-analysis. These tests control the distribution range of the dataset, and the heterogeneity and bias can be assessed using forest and funnel plots, respectively. The Meta package in R was employed to evaluate heterogeneity and bias.

The dataset was initially exported as a CSV file, with concentration data assigned to the experimental group and mean values to the control group. The Meta package was

executed in r4.3.0, generating experimental results based on descriptions of the forest and funnel plots. The forest plot outputs include heterogeneity, measured by I^2 , where I^2 exceeding 50% indicates strong heterogeneity and significant differences. A p value less than 0.01 indicates that the heterogeneity test falls within the 99% confidence interval. A funnel plot was used to detect sample bias, with a high or insufficient odds ratio indicating sample bias. Samples falling outside the inverted triangle were considered biased.

In this study, we divided the different regions of each indicator in the dataset. Due to space constraints, we did not list the distribution of a single sampling point but instead discussed the distribution of three groups: the west coast, Pearl River Delta, and East Coast. The forest map reveals spatial differences among the elemental compounds. With the exception of sulfate, all other elements showed a single effect mode, indicating similar sources in the three regions. Among them, Cd, Cr, Cu, Mn, and Pb showed the smallest differences ($p = 1$), reflecting consistent trends in elemental compounds across the three regions. Sulfate, on the other hand, showed a complex effect model, potentially influenced by regional factors. The time difference is illustrated in **Fig. 3** of the manuscript using a box diagram, which demonstrates spatial and temporal differences in each year.

Biased data is identified through a funnel diagram. This means that data points outside the funnel need to be screened out. This is necessary because data with excessive deviation can have a significant impact on the results of the model. During the analysis of the funnel diagram, we observed that the symmetry of metal elements and N and P elements is good, and the data sets are all within the funnel. Therefore, there is no need to screen out biased data in these cases. However, some data in C and S need to be excluded from the funnel diagram.

In this study, we hope that research bias will not affect the validity of data subsets used to extract evaluation indexes. Therefore, the most ideal state is that there is no heterogeneity and no bias in the data subsets after removing outliers.

Different from other studies of single effects through meta-analysis, we need to obtain a dataset of complex effects. That is, the samples need to show heterogeneity in

the forest map and pass the funnel plot. The reason for meeting these conditions is to ensure that the dataset used for calculating evaluation indexes not only captures the regional differences in different regions of the sampling sites but also does not excessively exaggerate the contribution of sample data distribution in a certain region to the mean values. For datasets with spatial heterogeneity and bias, it is necessary to further analyze the necessity of extracting specific regional evaluation indexes to obtain more accurate evaluation results.

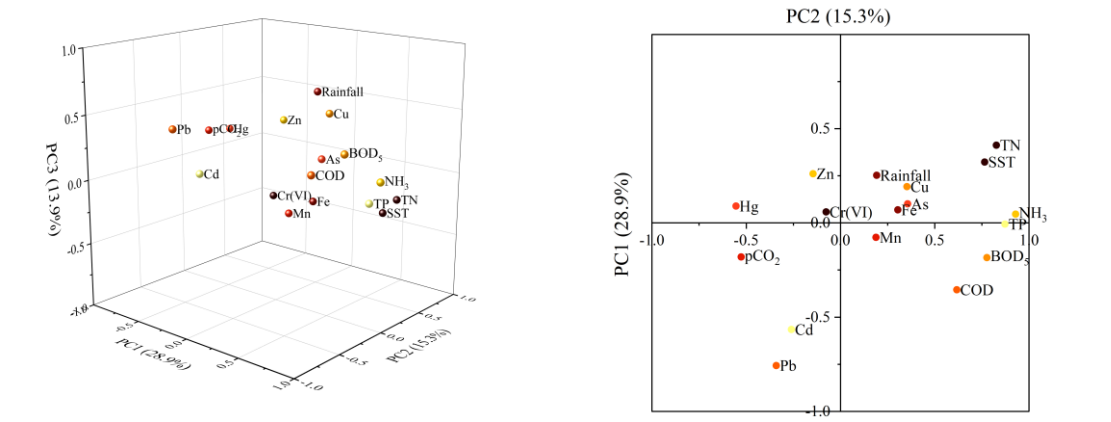
SB.2. Variable information of the model

The detailed information of the three variables used in the receptor model, namely sea surface temperature, rainfall flux, and abiotic surface aqueous partial pressure of CO₂ (*p*CO₂), can be found in Appendix I.

SB.3. Pre-test and analysis results of principal component analysis

Supplementary Table 3. Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett test of sphericity

KMO measure of sampling adequacy	Bartlett test of sphericity		
	Approximate chi-square	df	Sig
0.643	1291	210	0



Supplementary Fig. 4. PCA illustrating the relationship between estuarine pollutants and environmental factors.

Supplementary Table 4. Total variance explained by PCA

Principal Component Number	Eigenvalue	Percentage of Variance (%)	Cumulative (%)
1	6.075	28.927	28.927
2	3.210	15.284	44.211
3	2.976	13.934	58.145

Supplementary Table 5. Component matrix of PCA

	PC1	PC2	PC3
SST	0.765	0.323	−0.210
Rainfall	0.192	0.252	0.643
<i>p</i> CO ₂	−0.526	−0.181	0.312
Hg	−0.554	0.089	0.272
Pb	−0.340	−0.757	0.452
Cu	0.353	0.192	0.508
Zn	−0.145	0.261	0.373
Cd	−0.260	−0.566	0.093
Cr(VI)	−0.076	0.058	−0.178
Fe	0.304	0.069	−0.153
Mn	0.189	−0.077	−0.235
As	0.356	0.101	0.178
COD	0.617	−0.355	0.199
BOD	0.777	−0.184	0.343
NH ₃	0.929	0.046	0.118
TP	0.873	−0.007	−0.039
TN	0.827	0.412	−0.116

SB.4. Identification of principal component factors

In terms of component variable 1, DO, nitrogen, and phosphorus show high load capacity (**Supplementary Fig. 4**). Temperature affects aquatic biological activity, oxygen solubility, redox rate, and molecular diffusion rate, as well as metal release from sediments or suspended particles ³. Due to the limited content of nutrient elements in many inshore waters, the biogeochemistry and environmental hydrodynamics of sediments directly affect the nutrient dynamics of the estuary ⁴⁻⁶. The anoxic environment caused by high temperature in summer changes the path of nitrogen removal and recycling, disconnects nitrification-denitrification coupling and mineralization during anoxia, and promotes dissimilatory nitrate reduction to ammonium (DNRA), which will lead to the increase of NH₃ release from deposits ⁷. The effect of temperature on the fate behavior of metals in water is complex and uncertain. The increase of temperature accelerates the oxidation activity of sulfides, thus increasing the solubility of metals, thus increasing the transformation of metals from residual state to dissolved state ⁸.

Hg, Pb, and Cd have significant responses to the component variable 2 (**Supplementary Fig. 4**), the load of nutrients such as nitrogen and phosphorus on PC2 is very low, so the influence of domestic sewage input can be excluded. Hg, Pb, and Cd from vehicle exhaust, coal, petrol, and diesel fuel combustion can be transported through the atmosphere ⁹⁻¹¹, this is consistent with other studies ^{12,13}. We infer that the effect of atmospheric deposition contributes mainly to PC2. There is evidence that the importance of atmospheric Pb flux is 15–30 times that of rivers ¹⁴, and it has been determined that atmospheric deposition is the main source of Hg in the ocean ¹⁵, which means that atmospheric deposition is an important (or even major) contribution to offshore sediments.

Loads of pH, Fe, and Zn on component variable 3 are very high. It is speculated that PC3 may be closely related to the pH of seawater and may be CO₂ in the atmosphere. Due to global climate change, the concentration of CO₂ in the atmosphere increases. After absorbing CO₂, seawater shows a decrease in pH value and an increase

in HCO_3^- concentration, that is, ocean acidification^{16,17}. This will affect the state of trace metals in seawater from both thermodynamic and kinetic aspects¹⁸. Several metals exhibit different pH responses because changes in pH values in seawater strongly affect those metals that form strong complexes with hydroxides (OH^-) and carbonates (CO_3^{2-}). For example, acidification in the ocean increases the solubility of Fe(III) and promotes the supply of iron in water, which at the same time changes the form of environmental iron^{18,19}. In contrast, trace metals that exist as free ions or form strong complexes with chlorides do not respond violently to changes in pH. In addition, CO_2 -induced marine acidification can significantly affect the metabolism of nitrifying bacteria at the transcriptional level, thus inhibiting nitrification in estuarine and coastal waters²⁰, and reducing denitrification by interfering with the energy metabolism of fungi in sediments²¹⁻²⁶.

SB.5 Description of the receptor model

The data set from 2015 to 2022 was used in the testing phase. The effect of the model can be measured by the R^2 value, and the closer the R^2 value is to 1, the stronger the data interpretation ability of the model is. According to all the performance statistics, Hg became the best simulation index of the receptor model, followed by COD, and TN (Supplementary Table 6). However, the model has a weak ability to explain S^{2-} and SO_4^{2-} . There are differences in the explanatory ability of the model to the variability of different elements, which may be caused by the variation range of element concentration, the number of influencing factors, and the influencing factors complexity.

The mean absolute error (MAE) and mean square error (MSE) are commonly used to evaluate the accuracy of a model's predictions. Smaller MAE and MSE values indicate that the model's predictions are closer to the actual observed values. The evaluation results show that the MAE values for the different elements range from 0.000 to 63.300, while the MSE values range from 0.000 to 5537.958. Among them, BOD_5 has the highest MAE value of 69.337, and SO_4^{2-} has the highest MSE value of 5537.958, indicating significant errors in predicting BOD_5 and SO_4^{2-} elements. On the contrary, for the COD and $\text{NH}_3\text{-N}$ indicators, the model has smaller MAE and MSE

values, indicating higher accuracy in predicting these elements. Overall, the model shows variations in predicting different elements and lower explanatory power for certain complex factors. However, the model's output generally fits well with the long-term trends of trace elements and nutrient elements in the water body, especially for elements strongly influenced by the three variables. The model's accuracy and effect size meet the requirements for studying the long-term changes in seawater quality under climate change in the research area.

Supplementary Table 6. Effect size and accuracy of receptor models

Compo nent	Human activity background						R ²	MAE	MSE
	Coefficient		value (mg L ⁻¹)						
	SST (°C)	Rainfall (kg m ⁻² s ⁻¹)	pCO ₂ (Pa)	West Coast	PRD	CoreEast Coast			
Cr(VI)	1.16×10 ⁻⁴	-2.14	-2.29×10 ⁻⁵	2.26×10 ⁻³	2.20×10 ⁻³	2.00×10 ⁻³	0.890	0.000	0.000
Cu	1.75×10 ⁻⁴	-1.39×10	-2.68×10 ⁻⁵	2.87×10 ⁻³	3.35×10 ⁻³	3.09×10 ⁻³	0.369	0.001	0.000
Fe	4.63×10 ⁻³	5.62×10	-9.20×10 ⁻⁴	1.16×10 ⁻¹	8.78×10 ⁻²	6.17×10 ⁻²	0.573	0.003	0.000
Hg	8.20×10 ⁻⁷	-2.53×10 ⁻²	-9.60×10 ⁻⁸	2.09×10 ⁻⁵	1.37×10 ⁻⁵	1.87×10 ⁻⁵	0.840	0.086	0.008
Mn	2.89×10 ⁻³	-6.25×10	-1.42×10 ⁻³	2.08×10 ⁻²	2.69×10 ⁻²	2.44×10 ⁻²	0.370	0.000	0.000
Pb	6.66×10 ⁻⁴	-1.09×10	-2.62×10 ⁻⁴	5.11×10 ⁻⁴	1.98×10 ⁻⁴	5.65×10 ⁻⁴	0.500	0.021	0.001
Zn	5.22×10 ⁻⁶	-9.06×10 ⁻²	-2.19×10 ⁻⁶	1.55×10 ⁻²	5.45×10 ⁻³	5.95×10 ⁻³	0.467	0.000	0.000
Cd	7.59×10 ⁻⁵	4.35×10 ⁻¹	-1.83×10 ⁻⁵	4.47×10 ⁻⁵	3.93×10 ⁻⁵	7.64×10 ⁻⁵	0.444	0.007	0.000
As	1.16×10 ⁻⁴	-2.14	-2.29×10 ⁻⁵	3.33×10 ⁻⁴	1.56×10 ⁻³	1.39×10 ⁻³	0.670	0.008	0.001
COD	7.21×10 ⁻¹	-1.14×10 ⁴	-1.68×10 ⁻¹	1.53×10 ¹	1.16×10 ¹	1.21×10 ¹	0.767	1.230	2.601
BOD ₅	1.36×10 ⁻¹	-2.86×10 ³	-3.43×10 ⁻²	2.67×10 ⁰	2.17×10 ⁰	2.16×10 ⁰	0.721	6.672	69.337
NH ₃ -N	2.10×10 ⁻²	5.59×10 ²	-2.85×10 ⁻³	5.35×10 ¹	5.02×10 ¹	3.80×10 ¹	0.357	0.383	0.399
NO ₃ ⁻	7.23×10 ⁻²	-2.77×10 ³	-1.53×10 ⁻²	1.23×10 ⁰	1.69×10 ⁰	9.04×10 ¹	0.733	0.768	0.850
TN	1.23×10 ⁻¹	3.62×10 ³	-1.97×10 ⁻²	2.55×10 ⁰	2.96×10 ⁰	1.98×10 ⁰	0.770	1.026	2.115
TP	3.79×10 ⁻³	6.97×10	2.67×10 ⁻⁴	1.47×10 ¹	1.27×10 ¹	9.35×10 ²	0.611	0.065	0.011
S ²⁻	1.58×10 ⁻³	1.36×10 ¹	-8.95×10 ⁻⁴	6.18×10 ³	5.47×10 ³	1.27×10 ²	0.156	0.008	0.000
SO ₄ ²⁻	1.26×10	4.70×10 ⁴	7.16×10	4.55×10 ¹	6.68×10 ¹	6.89×10 ¹	0.323	63.300	5537.958

Part C Model verification

SC.1 Identify robustness in data subsets using variable substitution

In meta-analyze, the studies included in the analysis often possess distinct characteristics, and these differences may impact the overall effect size. Robustness analysis is an uncertainty analysis technique that examines the extent to which changes in relevant factors influence a specific key index or group of indicators through quantitative analysis. Its fundamental purpose is to elucidate how changes in these factors affect the key indicators by systematically altering the values of the relevant variables. By taking into account the characteristics of the included studies, robustness analysis can exclude low-quality studies or those utilizing different variables, subsequently conducting a combined analysis to compare with the original effect size prior to exclusion. This comparison enables exploration of the influence of the excluded studies on the overall effect size. Robustness analysis primarily focuses on comparing the combined effect size with the initial effect size derived from repeated meta-analyses. If the results exhibit minimal changes in effect size when variables are altered, it indicates relative robustness in the model's outcomes. Conversely, if significant differences or even contradictory conclusions emerge, it suggests poor robustness in the meta-analysis results, necessitating caution in interpreting and drawing conclusions from the findings.

Supplementary Table 7. Variables and replacement variables for robustness analysis

Variable	Associated variable	R^2	Sample size
SST	Dissolved oxygen	– 0.393**	2548
Rainfall flux	Total precipitation	0.498**	2549
Rainfall flux	Convective precipitation	0.543**	2549
$p\text{CO}_2$	pH	0.460**	2466

For improved verification, we established a total of 9 reference variables. Among these variables, dissolved oxygen was associated with temperature. The variables

linked to deposition were total precipitation after deviation correction and convective precipitation. pH was the variable associated with $p\text{CO}_2$. Additionally, surface runoff and conductivity were included in the control group with alternate fitting indexes. This information is displayed in Supplementary **Table 7**.

We used SST, sedimentation, and PCO_2 as fitting indicators, and tested for robustness with dissolved oxygen, convective precipitation, and pH as controls. It is used to investigate the effect size (R^2) of different indicators. The results are shown in Supplementary **Table 8**.

Supplementary Table 8. Fitting effect size R^2 of the experimental group and control group

Experimental group							
Cd	Cr	Cu	Fe	Hg	Mn	Pb	Zn
0.421	0.404	0.369	0.564	0.850	0.344	0.483	0.467
BOD	COD	$\text{NH}_3\text{-N}$	NO_3^-	TN	TP	S^{2-}	SO_4^{2-}
0.709	0.764	0.348	0.733	0.759	0.582	0.129	0.294
Control group							
Cd	Cr	Cu	Fe	Hg	Mn	Pb	Zn
0.412	0.407	0.375	0.557	0.843	0.478	0.479	0.476
BOD	COD	$\text{NH}_3\text{-N}$	NO_3^-	TN	TP	S^{2-}	SO_4^{2-}
0.733	0.771	0.450	0.728	0.794	0.657	0.203	0.305

In the comparison between the experimental group and the control group, we can see that R^2 has good coincidence, which proves that the correlation of the model is not caused by accidental factors and has reliability. Due to the variable requirements of the CMIP6 model, the process is to verify the robustness of the fit of the SST, wet sedimentation, and $p\text{CO}_2$ variables, rather than to discuss which variable is better fitted.

SC.2 Identify sensibility in data subsets using the sensitivity factor method

The modeling process involves making certain assumptions about the problem at hand. It is crucial to consider the sensitivity of the results to each assumption and

parameter. Sensitivity analysis is commonly performed based on assumptions made about the data. Expressing sensitivity data in terms of relative or percentage change is often more natural and applicable than using absolute change. A model is considered robust when the results derived from it remain correct even if the model itself is not completely accurate. Sensitivity analysis often involves identifying new constraints or determining the areas that require immediate attention for sensitivity analysis, based on some idealized assumptions of the model.

By decision, we will have SST, deposition, and $p\text{CO}_2$ as model variables, and we will discuss the response of different elements to these three variables, i.e., sensitivity analysis, which is calculated in Formula 2 of **manuscript 3.2.1**. In the output result, if the sensitivity of the variable is > 0.25 , it is sensitive to the indicator. In the output, we see that the sensitivity parameter is close to or reaches the requirement of 0.25, and there is no need to reduce variables to the model, and the above three variables will be used to describe different indicators.

Supplementary Table 9. Sensitivity of different indicators to three variables

	SST	Rainfall	$p\text{CO}_2$
Cd	0.30	0.88	0.28
Cr	0.14	0.40	0.13
Cu	0.66	0.57	0.44
Fe	0.31	0.34	0.31
Hg	0.81	0.33	0.49
Mn	0.33	0.97	0.31
Pb	0.29	0.67	0.36
Zn	0.83	0.28	0.79
BOD	0.31	0.60	0.40
COD	0.32	0.36	0.41
$\text{NH}_3\text{-N}$	0.26	0.77	0.25
NO_3^-	0.25	0.66	0.21
TN	0.33	0.96	0.31
TP	0.26	0.78	0.25
S^{2-}	0.72	0.27	0.68
SO_4^{2-}	0.82	0.26	0.79

Through the analysis of forest plots and funnel plots, the data subsets after removing outliers were validated and evaluated. The evaluation results showed that the

distribution of elements and compounds exhibited acceptable spatial differences, and the sources of elements in the three regions were generally similar. Biased data sets were identified and excluded through the funnel plots to ensure that the outliers did not affect the validity of the extracted data subsets while retaining the regional differences in the data. The results of the robustness analysis indicated that the model's results were consistent with changes in the variables, demonstrating the model's robustness. The sensitivity analysis revealed variations in the sensitivity of different indicators to SST, deposition, and $p\text{CO}_2$. These results provide a reliable data foundation for further research.

SC.3 Loss function values of three models

The output results of the selected machine learning models, Random Forest and XGBoost, will be presented in this support material. A comparative analysis between these two methods and the receptor model should be conducted prior to the presentation.

In machine learning, supervised learning is essentially a mapping relationship that aims to learn from a series of training samples, with the goal of accurately predicting the output for an unseen input x . The loss function, which maps the value of a random event or its related random variables to a non-negative real number, plays a crucial role in this process. It represents the "risk" or "loss" associated with the random event and is commonly employed for parametric estimation of model parameters. The loss function quantifies the difference between the model's output and the actual output, serving as a foundation for optimizing the model.

The Mean Squared Error (MSE) Loss, also known as L2 Loss, is widely used in machine learning and deep learning regression tasks. L2 refers to the L2 Norm, which represents the Euclidean Norm. By calculating the Euclidean distance between the predicted value and the true value, MSE provides a measure of the difference between them. Consequently, L2 Loss is derived. Its basic form can be expressed as follows,

$$J_{\text{MSE}} = \frac{1}{N} \sum_i^n (y_i - \hat{y}_i)^2 \quad (1)$$

Mean Absolute Error (MAE) is another commonly used loss function, also known as L1 Loss. The L1 norm refers to the sum of the absolute values of each element, so

we can also observe the connection between L1 Loss and MAE. Its basic form can be expressed as follows:

$$J_{MAE} = \frac{1}{N} \sum_i^n |y_i - \hat{y}_i| \quad (2)$$

The main distinction between MAE and MSE as loss functions is that MSE losses typically converge faster than MAE due to the distinction between Squaring and Linearity. However, MAE losses are more robust to outliers, meaning they are less affected by outlier noise. We have learned about the advantages and disadvantages of MSE and MAE losses: MSE losses converge quickly but are vulnerable to outlier noise, while MAE is more robust to outliers but converges slowly.

According to the methods and materials described in the manuscript, the relationship between the iron concentration and the properties of the solution is evaluated, and the influence ratio is used to quantify the influence of the physical properties on the temporal and spatial variation of the iron concentration. We use machine learning method for feature decomposition and list the loss function, R^2 , in a table to compare the accuracy of the three models. As shown in Supplementary Table 10.

Supplementary Table 10. Effect size (R^2), means absolute error loss (MAE), and mean square error loss (MSE) for the three models

	RandomForest			Xgboost			Receptor model		
	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE
Cd	0.661	0.000	0.000	Underfitting			0.444	0.000	0.000
As	0.451	0.001	0.000	0.635	0.000	0.000	0.670	0.000	0.000
Cr(VI)	0.625	0.001	0.000	0.449	0.001	0.000	0.890	0.001	0.000
Cu	0.494	0.002	0.000	0.894	0.001	0.000	0.369	0.003	0.000
Fe	0.649	0.039	0.002	0.996	0.004	0.000	0.573	0.086	0.008
Hg	0.431	0.000	0.000	Underfitting			0.840	0.000	0.000
Mn	0.802	0.009	0.001	0.993	0.001	0.000	0.370	0.021	0.001
Pb	0.484	0.000	0.000	0.251	0.000	0.000	0.500	0.000	0.000
Zn	0.506	0.004	0.000	0.941	0.001	0.000	0.467	0.007	0.000

	RandomForest			Xgboost			Receptor model		
	R^2	MAE	MSE	R^2	MAE	MSE	R^2	MAE	MSE
COD	0.455	4.022	26.082	0.979	0.696	1.018	0.767	3.622	21.649
BOD ₅	0.491	0.758	0.994	0.998	0.159	0.565	0.721	0.641	0.763
NH ₃	0.639	0.247	0.125	0.986	0.044	0.005	0.357	0.318	0.243
NO ₃ ⁻	0.473	0.429	0.296	0.982	0.068	0.010	0.733	0.380	0.237
TN	0.512	0.668	0.834	0.971	0.146	0.050	0.770	0.637	0.752
TP	0.590	0.043	0.003	0.971	0.011	0.000	0.611	0.036	0.003
S ²⁻	0.817	0.005	0.000	0.985	0.002	0.000	0.156	0.008	0.000
SO ₄ ²⁻	0.882	21.408	1091.396	0.999	0.590	0.828	0.323	63.300	5537.958

In comparing the three models, it is evident that both the receptor model and the random forest model have their respective advantages and disadvantages in terms of R^2 . The receptor model exhibits a higher R^2 in C and N compounds, but a lower R^2 in trace metals compared to the random forest model. On the other hand, **the R^2 of Xgboost is nearly 1.000, indicating a potential overfitting error.** When comparing MAE and MSE, the receptor model shows similar results to the random forest and Xgboost models, suggesting that the information loss function does not significantly increase. **Overall, based on the control experiment, the receptor model exhibited poor fitting for S²⁻ and SO₄²⁻, while the prediction for other elements remained within an acceptable range.**

SC.4 The fitting results of the random forest and Xgboost model

Random forest model

Analysis procedure:

1. Establish a random forest regression model using the training set data.
2. Calculate Feature importance by establishing random forests.
3. Apply the established random forest regression model to both the training and test data to obtain the Model evaluation results.
4. Due to the randomness of the random forest, the results may vary for each operation.

If the training model is saved, subsequent data can be directly uploaded and substituted into the training model for calculation and prediction.

5. Note: Random forests cannot provide a definite equation like traditional models. The model is typically evaluated by testing the accuracy of data predictions.

Algorithm configuration:

Algorithm: Random Forest regression

Variable: Variable X: {Rainfall, Runoff, Convective precipitation, pH, Temperature, Conductivity, $p\text{CO}_2$, DO, Wet deposition}

The training time: 0.385s

Data segmentation: 0.7

Data shuffle: No

Cross verify: No

Node split evaluation criteria: MSE

The maximum feature proportion considered in the partition: None

Minimum sample number of internal node splits: 2

The minimum sample number of leaf nodes: 1

The minimum weight of the sample in the leaf node: 0

Maximum depth of tree: 10

Maximum number of leaf nodes: 50

The threshold of node partition impurity: 0

Number of decision trees: 100

There is put back sampling: true

Out-of-pocket data test: false

XGBoost

Analysis procedure:

1. Build the XGBoost regression model using the training dataset.
2. Calculate the Feature importance by utilizing the established XGBoost model.
3. Apply the established XGBoost regression model to both the training and test datasets to obtain the evaluation results of the model.

4. Since XGBoost relies on randomness, the results may vary with each operation. If the training model is saved, subsequent data can be directly uploaded and used for calculation and prediction with the training model.

5. Note: XGBoost does not provide a definite equation like traditional models. The model is typically evaluated by testing the accuracy of the predictions on the data.

Algorithm configuration:

Algorithm: XGBoost regression

Variable: Variable X: {Rainfall, Runoff, Convective precipitation, pH, Temperature, Conductivity, $p\text{CO}_2$, DO, Wet deposition}

The training time: 0.794 seconds

Data segmentation: 0.7

Data shuffle: No

Cross verify: No

Base learner: gbtree

Number of base learners: 100

Learning rate: 0.1

L1 regular term: 0

L2 regular term: 1

Sample collection sampling rate: 1

Tree feature sampling rate: 1

Node feature sampling rate: 1

The minimum weight of the sample in the leaf node: 0

Maximum depth of tree: 10

Supplementary References:

- 1 Li, Y., & Niu, Z.G., China Water Cover Map (2020/10 m). V2. Science Data Bank,. <https://cstr.cn/31253.11.sciencedb.00880>. CSTR:31253.11.sciencedb.00880(2022).
- 2 GaryBikini, ChinaAdminDivisonSHP. v24.02.06, <https://doi.org/10.5281/zenodo.10624971> (2024).
- 3 Harmesa, Wahyudi, A. a. J., Lestari & Taufiqurrahman, E. Variability of trace metals in coastal and estuary: Distribution, profile, and drivers. *Marine Pollution Bulletin* **174**, 113173, doi:<https://doi.org/10.1016/j.marpolbul.2021.113173> (2022).
- 4 Rowe, G. T., Clifford, C. H., Smith, K. L. & Hamilton, P. L. Benthic nutrient regeneration and its coupling to primary productivity in coastal waters. *Nature* **255**, 215-217, doi:10.1038/255215a0 (1975).
- 5 Soetaert, K. & Middelburg, J. J. Modeling eutrophication and oligotrophication of shallow-water marine systems: the importance of sediments under stratified and well-mixed conditions. *Hydrobiologia* **629**, 239-254, doi:10.1007/s10750-009-9777-x (2009).
- 6 Lorke, A., Müller, B., Maerki, M. & Wüest, A. Breathing sediments: The control of diffusive transport across the sediment—water interface by periodic boundary-layer turbulence. *Limnology and Oceanography* **48**, 2077-2085, doi:<https://doi.org/10.4319/lo.2003.48.6.2077> (2003).
- 7 Roberts, K. L., Eate, V. M., Eyre, B. D., Holland, D. P. & Cook, P. L. M. Hypoxic events stimulate nitrogen recycling in a shallow salt-wedge estuary: The Yarra River estuary, Australia. *Limnology and Oceanography* **57**, 1427-1442, doi:<https://doi.org/10.4319/lo.2012.57.5.1427> (2012).
- 8 Fu, S. & Lu, J. Temperature-driven variation in the removal of heavy metals from contaminated tailings leaching in northern Norway. *Environmental Monitoring and Assessment* **191**, 123, doi:10.1007/s10661-019-7244-3 (2019).
- 9 Adokoh, C. K. *et al.* Statistical Evaluation of Environmental Contamination, Distribution and Source Assessment of Heavy Metals (Aluminum, Arsenic, Cadmium, and Mercury) in Some Lagoons and an Estuary Along the Coastal Belt of Ghana. *Archives of Environmental Contamination and Toxicology* **61**, 389-400, doi:10.1007/s00244-011-9643-5 (2011).
- 10 Lin, Y., Vogt, R. & Larssen, T. Environmental mercury in China: A review. *Environmental Toxicology and Chemistry* **31**, 2431-2444, doi:10.1002/etc.1980 (2012).
- 11 Burgos-Núñez, S., Navarro-Frómata, A., Marrugo-Negrete, J., Enamorado-Montes, G. & Urango-Cárdenas, I. Polycyclic aromatic hydrocarbons and heavy metals in the Cispata Bay, Colombia: A marine tropical ecosystem. *Marine Pollution Bulletin* **120**, 379-386, doi:<https://doi.org/10.1016/j.marpolbul.2017.05.016> (2017).
- 12 Rose, N. L. & Rippey, B. The historical record of PAH, PCB, trace metal and fly-ash particle deposition at a remote lake in north-west Scotland. *Environmental Pollution* **117**, 121-132, doi:[https://doi.org/10.1016/S0269-7491\(01\)00149-X](https://doi.org/10.1016/S0269-7491(01)00149-X) (2002).
- 13 Rippey, B. *et al.* An assessment of toxicity in profundal lake sediment due to deposition of heavy metals and persistent organic pollutants from the atmosphere. *Environment International* **34**, 345-356, doi:<https://doi.org/10.1016/j.envint.2007.03.006> (2008).

- 14 Guerzoni, S. *et al.* The role of atmospheric deposition in the biogeochemistry of the Mediterranean Sea. *Progress in Oceanography* **44**, 147-190, doi:[https://doi.org/10.1016/S0079-6611\(99\)00024-5](https://doi.org/10.1016/S0079-6611(99)00024-5) (1999).
- 15 Mason, R. P. & Sheu, G. R. Role of the ocean in the global mercury cycle. *Global Biogeochemical Cycles* **16**, 40-41-40-14, doi:<https://doi.org/10.1029/2001GB001440> (2002).
16 <14.pdf>. doi:10.3354/meps10082.
- 17 Henderson, G. M., Achterberg, E. P. & Bopp, L. Changing Trace Element Cycles in the 21st Century Ocean. *Elements* **14**, 409-413, doi:10.2138/gselements.14.6.409 (2018).
- 18 Zhu, K. *et al.* Equilibrium calculations of iron speciation and apparent iron solubility in the Celtic Sea at ambient seawater pH using the NICA-Donnan model. *Marine Chemistry* **237**, 104038, doi:<https://doi.org/10.1016/j.marchem.2021.104038> (2021).
- 19 Zhu, K. *et al.* Influence of Changes in pH and Temperature on the Distribution of Apparent Iron Solubility in the Oceans. *Global Biogeochemical Cycles* **37**, e2022GB007617, doi:<https://doi.org/10.1029/2022GB007617> (2023).
- 20 Zhou, J. *et al.* Effects of acidification on nitrification and associated nitrous oxide emission in estuarine and coastal waters. *Nature Communications* **14**, 1380, doi:10.1038/s41467-023-37104-9 (2023).
- 21 Su, X. *et al.* Denitrification and N₂O Emission in Estuarine Sediments in Response to Ocean Acidification: From Process to Mechanism. *Environmental Science & Technology* **56**, 14828-14839, doi:10.1021/acs.est.2c03550 (2022).
- 22 Zhu, H., Bing, H., Yi, H., Wu, Y. & Sun, Z. Spatial Distribution and Contamination Assessment of Heavy Metals in Surface Sediments of the Caofeidian Adjacent Sea after the Land Reclamation, Bohai Bay. *Journal of Chemistry* **2018**, 2049353, doi:<https://doi.org/10.1155/2018/2049353> (2018).
- 23 Zhang, J. *et al.* Spatial distribution and correlation characteristics of heavy metals in the seawater, suspended particulate matter and sediments in Zhanjiang Bay, China. *PLOS ONE* **13**, e0201414, doi:10.1371/journal.pone.0201414 (2018).
- 24 Zhang, C. *et al.* Effects of sediment geochemical properties on heavy metal bioavailability. *Environment International* **73**, 270-281, doi:<https://doi.org/10.1016/j.envint.2014.08.010> (2014).
- 25 Mellett, T. & Buck, K. N. Spatial and temporal variability of trace metals (Fe, Cu, Mn, Zn, Co, Ni, Cd, Pb), iron and copper speciation, and electroactive Fe-binding humic substances in surface waters of the eastern Gulf of Mexico. *Marine Chemistry* **227**, 103891, doi:<https://doi.org/10.1016/j.marchem.2020.103891> (2020).
- 26 Herbert, E. R. *et al.* A global perspective on wetland salinization: ecological consequences of a growing threat to freshwater wetlands. *Ecosphere* **6**, art206, doi:<https://doi.org/10.1890/ES14-00534.1> (2015).

Appendix I

(1) Variables used for sea surface temperature

Global Attributes:

Conventions	= 'CF-1.7 CMIP-6.2'
activity_id	= 'ScenarioMIP AerChemMIP'
branch_method	= 'Standard'
branch_time_in_child	= 0
branch_time_in_parent	= 2015
contact	= 'Dr. Xinyao Rong (rongxy@cma.gov.cn)'
creation_date	= '2019-06-11T05:58:01Z'
data_specs_version	= '01.00.30'
description	= 'SSP3-7.0'
experiment	= 'gap-filling scenario reaching 7.0 based on SSP3'
experiment_id	= 'ssp370'
external_variables	= 'areacello'
forcing_index	= 1
frequency	= 'mon'
further_info_url	= 'https://furtherinfo.es-doc.org/CMIP6.CAMS.CAMS- CSM1-0.ssp370.none.r1i1p1f1'
grid	= 'native ocean tri-polar grid'
grid_label	= 'gn'
history	= '2019-06-11T05:58:01Z ;rewrote data to be consistent with ScenarioMIP for variable tos found in table Omon.;
Output from CAMS-CSM1-0'	
initialization_index	= 1
institution	= 'Chinese Academy of Meteorological Sciences, Beijing 100081, China'
institution_id	= 'CAMS'
mip_era	= 'CMIP6'
nominal_resolution	= '100 km'
parent_activity_id	= 'CMIP'
parent_experiment_id	= 'historical'
parent_mip_era	= 'CMIP6'
parent_source_id	= 'CAMS-CSM1-0'
parent_time_units	= 'days since 2015-01-01'
parent_variant_label	= 'r1i1p1f1'
physics_index	= 1
product	= 'model-output'
realization_index	= 1
realm	= 'ocean'
references	= 'Model described by Rong et al (J. Meteor. Res., 2019, 576-591).'
run_variant	= 'forcing: greenhouse gases,aerosol,solar

constant,volcano,ozone'

source = 'CAMS_CSM 1.0 (2016):
aerosol: none
atmos: ECHAM5_CAMS (T106; 320 x 160
longitude/latitude; 31 levels; top level 10 mb)
atmosChem: none
land: CoLM 1.0
landIce: none
ocean: MOM4 (tripolar; 360 x 200 longitude/latitude,
primarily 1deg latitude/longitude, down to 1/3deg within 30deg of the equatorial tropics; 50 levels;
top grid cell 0-10 m)

ocnBgchem: none
seaIce: SIS 1.0'

source_id = 'CAMS-CSM1-0'
source_type = 'AOGCM'
sub_experiment = 'none'
sub_experiment_id = 'none'
table_id = 'Omon'
table_info = 'Creation Date:(30 April 2019)

MD5:cc2ae51c23960ce127b4d68ab3d8ac87'

title = 'CAMS-CSM1-0 output prepared for CMIP6'
tracking_id = 'hdl:21.14100/3b33f3fc-9159-424a-9502-3dbe2461eea7'
variable_id = 'tos'
variant_label = 'r1i1p1f1'
license = 'CMIP6 model data produced by Lawrence Livermore

PCMDI is licensed under a Creative Commons Attribution ShareAlike 4.0 International License (<https://creativecommons.org/licenses>). Consult <https://pcmdi.llnl.gov/CMIP6/TermsOfUse> for terms of use governing CMIP6 output, including citation requirements and proper acknowledgment. Further information about this data, including some limitations, can be found via the further_info_url (recorded as a global attribute in this file) and at <https://pcmdi.llnl.gov/>. The data producers and data providers make no warranty, either express or implied, including, but not limited to, warranties of merchantability and fitness for a particular purpose. All liabilities arising from the supply of the information (including any liability arising in negligence) are excluded to the fullest extent permitted by law.'

cmor_version = '3.4.0'

Dimensions:

time = 1020 (UNLIMITED)
j = 200
i = 360
bnds = 2
vertices = 4

Variables:

time

Size: 1020x1

Dimensions: time
Datatype: double
Attributes:
 bounds = 'time_bnds'
 units = 'days since 2015-1-1 00:00'
 calendar = '365_day'
 axis = 'T'
 long_name = 'time'
 standard_name = 'time'

time_bnds

Size: 2x1020
Dimensions: bnds,time
Datatype: double

j

Size: 200x1
Dimensions: j
Datatype: int32
Attributes:
 units = '1'
 long_name = 'cell index along second dimension'

i

Size: 360x1
Dimensions: i
Datatype: int32
Attributes:
 units = '1'
 long_name = 'cell index along first dimension'

latitude

Size: 360x200
Dimensions: i,j
Datatype: single
Attributes:
 standard_name = 'latitude'
 long_name = 'latitude'
 units = 'degrees_north'
 missing_value = 1.000000020040877e+20
 _FillValue = 1.000000020040877e+20
 bounds = 'vertices_latitude'

longitude

Size: 360x200
Dimensions: i,j
Datatype: single
Attributes:
 standard_name = 'longitude'

```

        long_name      = 'longitude'
        units           = 'degrees_east'
        missing_value = 1.000000020040877e+20
        _FillValue      = 1.000000020040877e+20
        bounds          = 'vertices_longitude'

vertices_latitude
    Size:          4x360x200
    Dimensions: vertices,i,j
    Datatype:      single
    Attributes:
        units       = 'degrees_north'
        missing_value = 1.000000020040877e+20
        _FillValue  = 1.000000020040877e+20

vertices_longitude
    Size:          4x360x200
    Dimensions: vertices,i,j
    Datatype:      single
    Attributes:
        units       = 'degrees_east'
        missing_value = 1.000000020040877e+20
        _FillValue  = 1.000000020040877e+20

tos
    Size:          360x200x1020
    Dimensions: i,j,time
    Datatype:      single
    Attributes:
        standard_name = 'sea_surface_temperature'
        long_name      = 'Sea Surface Temperature'
        comment        = 'Temperature of upper boundary of the liquid
ocean, including temperatures below sea-ice and floating ice shelves.'
        units          = 'degC'
        original_name  = 'temp'
        cell_methods   = 'area: mean where sea time: mean'
        cell_measures  = 'area: areacello'
        history        = '2019-06-11T05:57:56Z altered by CMOR: replaced
missing value flag (-1e+10) with standard missing value (1e+20).'
        missing_value = 1.000000020040877e+20
        _FillValue    = 1.000000020040877e+20
        coordinates    = 'latitude longitude'

```

(2) Variables used for wet deposition

Global Attributes:

```
CCCma_model_hash =
'1f91f92cb6d607391f44831504025d32fc44faa1'
CCCma_parent_runid = 'rc3.1-his01'
CCCma_pycmor_hash =
'33c30511acc319a98240633965a04ca99c26427e'
CCCma_runid = 'rc3.1-s7001'
Conventions = 'CF-1.7 CMIP-6.2'
YMDH_branch_time_in_child = '2015:01:01:00'
YMDH_branch_time_in_parent = '2015:01:01:00'
activity_id = 'ScenarioMIP AerChemMIP'
branch_method = 'Spin-up documentation'
branch_time_in_child = 60225
branch_time_in_parent = 60225
contact = 'ec.cccma.info-info.ccmac.ec@canada.ca'
creation_date = '2019-05-02T03:14:41Z'
data_specs_version = '01.00.29'
experiment = 'gap-filling scenario reaching 7.0 based on SSP3'
experiment_id = 'ssp370'
external_variables = 'areacello'
forcing_index = 1
frequency = 'mon'
further_info_url = 'https://furtherinfo.es-
doc.org/CMIP6.CCCma.CanESM5.ssp370.none.r1i1p1f1'
grid = 'ORCA1 tripolar grid, 1 deg with refinement to 1/3
deg within 20 degrees of the equator; 361 x 290 longitude/latitude; 45 vertical levels; top grid cell
0-6.19 m'
grid_label = 'gn'
history = '2019-05-02T03:14:41Z ;rewrote data to be
consistent with ScenarioMIP for variable prra found in table Omon.;
```

Output from \$runid'

```
initialization_index = 1
institution = 'Canadian Centre for Climate Modelling and Analysis,
Environment and Climate Change Canada, Victoria, BC V8P 5C2, Canada'
institution_id = 'CCCma'
mip_era = 'CMIP6'
nominal_resolution = '100 km'
parent_activity_id = 'CMIP'
parent_experiment_id = 'historical'
parent_mip_era = 'CMIP6'
parent_source_id = 'CanESM5'
parent_time_units = 'days since 1850-01-01 0:0:0.0'
parent_variant_label = 'r1i1p1f1'
```

physics_index	= 1
product	= 'model-output'
realization_index	= 1
realm	= 'atmos'
references	= 'Geophysical Model Development Special issue on CanESM5 (https://www.geosci-model-dev.net/special_issues.html)'
source	= 'CanESM5 (2019): aerosol: interactive atmos: CanAM5 (T63L49 native atmosphere, T63 Linear Gaussian Grid; 128 x 64 longitude/latitude; 49 levels; top level 1 hPa) atmosChem: specified oxidants for aerosols land: CLASS3.6/CTEM1.2 landIce: specified ice sheets ocean: NEMO3.4.1 (ORCA1 tripolar grid, 1 deg with refinement to 1/3 deg within 20 degrees of the equator; 361 x 290 longitude/latitude; 45 vertical levels; top grid cell 0-6.19 m) ocnBgchem: Canadian Model of Ocean Carbon (CMOC); NPZD ecosystem with OMIP prescribed carbonate chemistry seaIce: LIM2'
source_id	= 'CanESM5'
source_type	= 'AOGCM'
sub_experiment	= 'none'
sub_experiment_id	= 'none'
table_id	= 'Omon'
table_info	= 'Creation Date:(20 February 2019) MD5:374fbe5a2bcca535c40f7f23da271e49'
title	= 'CanESM5 output prepared for CMIP6'
variable_id	= 'prra'
variant_label	= 'r1i1p1f1'
version	= 'v20190429'
license	= 'CMIP6 model data produced by The Government of Canada (Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada) is licensed under a Creative Commons Attribution ShareAlike 4.0 International License (https://creativecommons.org/licenses). Consult https://pcmdi.llnl.gov/CMIP6/TermsOfUse for terms of use governing CMIP6 output, including citation requirements and proper acknowledgment. Further information about this data, including some limitations, can be found via the further_info_url (recorded as a global attribute in this file) and at https://pcmdi.llnl.gov/ . The data producers and data providers make no warranty, either express or implied, including, but not limited to, warranties of merchantability and fitness for a particular purpose. All liabilities arising from the supply of the information (including any liability arising in negligence) are excluded to the fullest extent permitted by law.'
cmor_version	= '3.4.0'
tracking_id	= 'hdl:21.14100/5a1e98a9-aa21-4f43-9b04- 4b8bad7aa563'

Dimensions:

time = 1032 (UNLIMITED)
bnds = 2
j = 291
i = 360
vertices = 4

Variables:

time

Size: 1032x1
Dimensions: time
Datatype: double
Attributes:
 bounds = 'time_bnds'
 units = 'days since 1850-01-01 0:0:0.0'
 calendar = '365_day'
 axis = 'T'
 long_name = 'time'
 standard_name = 'time'

time_bnds

Size: 2x1032
Dimensions: bnds,time
Datatype: double

j

Size: 291x1
Dimensions: j
Datatype: int32
Attributes:
 units = '1'
 long_name = 'cell index along second dimension'

i

Size: 360x1
Dimensions: i
Datatype: int32
Attributes:
 units = '1'
 long_name = 'cell index along first dimension'

latitude

Size: 360x291
Dimensions: i,j
Datatype: double
Attributes:
 standard_name = 'latitude'
 long_name = 'latitude'
 units = 'degrees_north'

```

        missing_value = 1e+20
        _FillValue     = 1e+20
        bounds         = 'vertices_latitude'

longitude
    Size:          360x291
    Dimensions:    i,j
    Datatype:      double
    Attributes:
        standard_name = 'longitude'
        long_name      = 'longitude'
        units          = 'degrees_east'
        missing_value = 1e+20
        _FillValue     = 1e+20
        bounds         = 'vertices_longitude'

vertices_latitude
    Size:          4x360x291
    Dimensions:    vertices,i,j
    Datatype:      double
    Attributes:
        units          = 'degrees_north'
        missing_value = 1e+20
        _FillValue     = 1e+20

vertices_longitude
    Size:          4x360x291
    Dimensions:    vertices,i,j
    Datatype:      double
    Attributes:
        units          = 'degrees_east'
        missing_value = 1e+20
        _FillValue     = 1e+20

prra
    Size:          360x291x1032
    Dimensions:    i,j,time
    Datatype:      single
    Attributes:
        standard_name = 'rainfall_flux'
        long_name      = 'Rainfall Flux where Ice Free Ocean over Sea'
        comment        = 'CanNEMO employs a virtual salt flux, but water
flux is used in the free surface formulation, and hence useful to report. This field includes rain falling
on sea ice as seen by CanNEMO., ScenarioMIP_table_comment: In accordance with common usage
in geophysical disciplines, 'flux' implies per unit area, called 'flux density' in physics.'
        units          = 'kg m-2 s-1'
        original_name = 'prlq'
        cell_methods   = 'area: mean where ice_free_sea over sea time: mean'

```

```
cell_measures = 'area: areacello'  
missing_value = 1.000000020040877e+20  
_FillValue     = 1.000000020040877e+20  
coordinates    = 'latitude longitude'
```

(3) Variables used for partial pressure of carbon dioxide

Global Attributes:

```
CCCma_model_hash =
'1f91f92cb6d607391f44831504025d32fc44faa1'
CCCma_parent_runid = 'rc3.1-his01'
CCCma_pycmor_hash =
'33c30511acc319a98240633965a04ca99c26427e'
CCCma_runid = 'rc3.1-s7001'
Conventions = 'CF-1.7 CMIP-6.2'
YMDH_branch_time_in_child = '2015:01:01:00'
YMDH_branch_time_in_parent = '2015:01:01:00'
activity_id = 'ScenarioMIP AerChemMIP'
branch_method = 'Spin-up documentation'
branch_time_in_child = 60225
branch_time_in_parent = 60225
contact = 'ec.cccma.info-info.ccmac.ec@canada.ca'
creation_date = '2019-05-02T03:15:11Z'
data_specs_version = '01.00.29'
experiment = 'gap-filling scenario reaching 7.0 based on SSP3'
experiment_id = 'ssp370'
external_variables = 'areacello'
forcing_index = 1
frequency = 'mon'
further_info_url = 'https://furtherinfo.es-
doc.org/CMIP6.CCCma.CanESM5.ssp370.none.r1i1p1f1'
grid = 'ORCA1 tripolar grid, 1 deg with refinement to 1/3
deg within 20 degrees of the equator; 361 x 290 longitude/latitude; 45 vertical levels; top grid cell
0-6.19 m'
grid_label = 'gn'
history = '2019-05-02T03:15:11Z ;rewrote data to be
consistent with ScenarioMIP for variable spco2abio found in table Omon.;
Output from $runid'
initialization_index = 1
institution = 'Canadian Centre for Climate Modelling and Analysis,
Environment and Climate Change Canada, Victoria, BC V8P 5C2, Canada'
institution_id = 'CCCma'
mip_era = 'CMIP6'
nominal_resolution = '100 km'
parent_activity_id = 'CMIP'
parent_experiment_id = 'historical'
parent_mip_era = 'CMIP6'
parent_source_id = 'CanESM5'
parent_time_units = 'days since 1850-01-01 0:0:0.0'
parent_variant_label = 'r1i1p1f1'
```

physics_index	= 1
product	= 'model-output'
realization_index	= 1
realm	= 'ocnBgchem'
references	= 'Geophysical Model Development Special issue on CanESM5 (https://www.geosci-model-dev.net/special_issues.html)'
source	= 'CanESM5 (2019): aerosol: interactive atmos: CanAM5 (T63L49 native atmosphere, T63 Linear Gaussian Grid; 128 x 64 longitude/latitude; 49 levels; top level 1 hPa) atmosChem: specified oxidants for aerosols land: CLASS3.6/CTEM1.2 landIce: specified ice sheets ocean: NEMO3.4.1 (ORCA1 tripolar grid, 1 deg with refinement to 1/3 deg within 20 degrees of the equator; 361 x 290 longitude/latitude; 45 vertical levels; top grid cell 0-6.19 m) ocnBgchem: Canadian Model of Ocean Carbon (CMOC); NPZD ecosystem with OMIP prescribed carbonate chemistry seaIce: LIM2'
source_id	= 'CanESM5'
source_type	= 'AOGCM'
sub_experiment	= 'none'
sub_experiment_id	= 'none'
table_id	= 'Omon'
table_info	= 'Creation Date:(20 February 2019) MD5:374fbe5a2bcca535c40f7f23da271e49'
title	= 'CanESM5 output prepared for CMIP6'
variable_id	= 'spco2abio'
variant_label	= 'r1i1p1f1'
version	= 'v20190429'
license	= 'CMIP6 model data produced by The Government of Canada (Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada) is licensed under a Creative Commons Attribution ShareAlike 4.0 International License (https://creativecommons.org/licenses). Consult https://pcmdi.llnl.gov/CMIP6/TermsOfUse for terms of use governing CMIP6 output, including citation requirements and proper acknowledgment. Further information about this data, including some limitations, can be found via the further_info_url (recorded as a global attribute in this file) and at https://pcmdi.llnl.gov/ . The data producers and data providers make no warranty, either express or implied, including, but not limited to, warranties of merchantability and fitness for a particular purpose. All liabilities arising from the supply of the information (including any liability arising in negligence) are excluded to the fullest extent permitted by law.'
cmor_version	= '3.4.0'
tracking_id	= 'hdl:21.14100/5f30979d-9927-4f27-9bdd- 655f40af474a'

Dimensions:

time = 1032 (UNLIMITED)
bnds = 2
j = 291
i = 360
vertices = 4

Variables:

time

Size: 1032x1
Dimensions: time
Datatype: double
Attributes:
 bounds = 'time_bnds'
 units = 'days since 1850-01-01 0:0:0.0'
 calendar = '365_day'
 axis = 'T'
 long_name = 'time'
 standard_name = 'time'

time_bnds

Size: 2x1032
Dimensions: bnds,time
Datatype: double

j

Size: 291x1
Dimensions: j
Datatype: int32
Attributes:
 units = '1'
 long_name = 'cell index along second dimension'

i

Size: 360x1
Dimensions: i
Datatype: int32
Attributes:
 units = '1'
 long_name = 'cell index along first dimension'

latitude

Size: 360x291
Dimensions: i,j
Datatype: double
Attributes:
 standard_name = 'latitude'
 long_name = 'latitude'
 units = 'degrees_north'

```

missing_value = 1e+20
_FillValue     = 1e+20
bounds         = 'vertices_latitude'

longitude
  Size:          360x291
  Dimensions: i,j
  Datatype:     double
  Attributes:
    standard_name = 'longitude'
    long_name     = 'longitude'
    units         = 'degrees_east'
    missing_value = 1e+20
    _FillValue    = 1e+20
    bounds       = 'vertices_longitude'

vertices_latitude
  Size:          4x360x291
  Dimensions: vertices,i,j
  Datatype:     double
  Attributes:
    units         = 'degrees_north'
    missing_value = 1e+20
    _FillValue    = 1e+20

vertices_longitude
  Size:          4x360x291
  Dimensions: vertices,i,j
  Datatype:     double
  Attributes:
    units         = 'degrees_east'
    missing_value = 1e+20
    _FillValue    = 1e+20

depth
  Size:          1x1
  Dimensions:
  Datatype:     double
  Attributes:
    units         = 'm'
    axis          = 'Z'
    positive      = 'down'
    long_name     = 'depth'
    standard_name = 'depth'

spco2abio
  Size:          360x291x1032
  Dimensions: i,j,time
  Datatype:     single

```

Attributes:

```
standard_name =  
'surface_carbon_dioxide_abiotic_analogue_partial_pressure_difference_between_sea_water_and_  
air'  
  
long_name = 'Abiotic Surface Aqueous Partial Pressure of CO2'  
comment = 'The surface called 'surface' means the lower  
boundary of the atmosphere. The chemical formula for carbon dioxide is CO2. In ocean  
biogeochemistry models, an 'abiotic analogue' is used to simulate the effect on a modelled variable  
when biological effects on ocean carbon concentration and alkalinity are ignored. The partial  
pressure of a gaseous constituent of air is the pressure which it alone would exert with unchanged  
temperature and number of moles per unit volume. The partial pressure of a dissolved gas in sea  
water is the partial pressure in air with which it would be in equilibrium. The partial pressure  
difference between sea water and air is positive when the partial pressure of the dissolved gas in sea  
water is greater than the partial pressure in air.'  
  
units = 'Pa'  
original_name = 'spco2a'  
history = 'mltby0.101325 2019-05-02T03:15:11Z altered by  
CMOR: Treated scalar dimension: 'depth'.  
  
cell_methods = 'area: mean where sea time: mean'  
cell_measures = 'area: areacello'  
coordinates = 'depth latitude longitude'  
missing_value = 1.000000020040877e+20  
_FillValue = 1.000000020040877e+20
```