



# CMEMS-LSCE: A global 0.25-degree, monthly reconstruction of the surface ocean carbonate system

Thi-Tuyet-Trang Chau<sup>1</sup>, Marion Gehlen<sup>1</sup>, Nicolas Metzler<sup>2</sup>, and Frédéric Chevallier<sup>1</sup>

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

<sup>2</sup>Laboratoire LOCEAN (IPSL), Sorbonne Université, CNRS-IRD-MNHN, Paris, F-75005, France

**Correspondence:** Thi-Tuyet-Trang CHAU (trang.chau@lscce.ipsl.fr, thi.tuyet.trang.chau@gmail.com)

**Abstract.** Observation-based data reconstructions of global surface ocean carbonate system variables play an essential role in monitoring the recent status of ocean carbon uptake and ocean acidification as well as their impacts on marine organisms and ecosystems. So far ongoing efforts are directed towards exploring new approaches to describe the complete marine carbonate system and to better recover its fine-scale features. In this respect, our research activities within the Copernicus Marine Environment Monitoring Service (CMEMS) aim at developing a sustainable production chain of observation-derived global ocean carbonate system datasets at high space-time resolution. As the start of the long-term objective, this study introduces a new global 0.25° monthly reconstruction, namely CMEMS-LSCE, for the period 1985-2021. The CMEMS-LSCE reconstruction derives datasets of six carbonate system variables including surface ocean partial pressure of CO<sub>2</sub> ( $p\text{CO}_2$ ), total alkalinity ( $A_T$ ), total dissolved inorganic carbon (DIC), surface ocean pH, and saturation states with respect to aragonite ( $\Omega_{ar}$ ) and calcite ( $\Omega_{ca}$ ). Reconstructing  $p\text{CO}_2$  relies on an ensemble of neural network models mapping gridded observation-based data provided by the Surface Ocean CO<sub>2</sub> Atlas (SOCAT). Surface ocean  $A_T$  is estimated with a multiple linear regression approach, and the remaining carbonate variables are resolved by CO<sub>2</sub> system speciation given the reconstructed  $p\text{CO}_2$  and  $A_T$ .  $1\sigma$ -uncertainty associated with these estimates is also provided. Here,  $\sigma$  stands for either ensemble standard deviation of  $p\text{CO}_2$  estimates or total uncertainty for each of the five other variables propagated through the processing chain with input data uncertainty. We demonstrate that the 0.25°-resolution  $p\text{CO}_2$  product outperforms a coarser spatial resolution (1°) thanks to a higher data coverage nearshore and a better description of horizontal and temporal variations in  $p\text{CO}_2$  across diverse ocean basins, particularly in the coastal-open-ocean continuum. Product qualification with observation-based data confirms reliable reconstructions with root-of-mean-square-deviation from observations less than 8%, 4%, and 1% relative to the global mean of  $p\text{CO}_2$ ,  $A_T$  (DIC), and pH. The global average  $1\sigma$ -uncertainty is below 5% and 8% for  $p\text{CO}_2$  and  $\Omega_{ar}$  ( $\Omega_{ca}$ ), 2% for  $A_T$  and DIC, and 0.4% for pH relative to their global mean values. Both model-observation misfit and model uncertainty indicate that coastal data reproduction still needs further improvement, wherein high temporal and horizontal gradients of carbonate variables and representative uncertainty from data sampling would be taken into account in priority. This study also presents a potential use case of the CMEMS-LSCE carbonate data product in tracking the recent state of ocean acidification.



## 1 Introduction

25 Between 1750 and 2019, the ocean took up an estimated 25% (or  $170 \pm 20$  PgC) of total cumulated anthropogenic  $\text{CO}_2$  ( $685 \pm 75$  PgC) emitted to the atmosphere (IPCC AR6 - the Sixth Assessment Report of the United Nations Intergovernmental Panel on Climate Change, Canadell et al., 2021). While the uptake of anthropogenic  $\text{CO}_2$  mitigates global warming it also profoundly modifies seawater chemistry in a suite of well-understood reactions (Orr et al., 2005) leading to an increase in hydrogen ion concentration ( $[\text{H}^+]$ ), as well as a decrease in carbonate ion concentration ( $[\text{CO}_3^{2-}]$ ) and in the saturation state of seawater ( $\Omega$ )  
30 with respect to calcium carbonate minerals ( $\text{CaCO}_3$ ). The increase in hydrogen ion concentration ( $[\text{H}^+]$ ) is commonly reported as a decrease in  $\text{pH}$  ( $\text{pH} = -\log[\text{H}^+]$ ) and referred to as ocean acidification.

Changes in carbonate chemistry impact calcifying plankton and benthos as a direct result of decreasing seawater saturation state with respect to  $\text{CaCO}_3$  (Fabry et al., 2008; Thomsen et al., 2015). Ocean acidification also modifies the production of marine trace gases exchanged at the air-sea interface (Hopkins et al., 2020), the availability of nutrients fueling primary  
35 production (Doney et al., 2009), as well as the speciation of pollutants (Millero et al., 2009; Hoffmann et al., 2012). These chemical changes interact with warming and ocean deoxygenation to drive major changes in marine ecosystems (Doney et al., 2020) and to alter global biogeochemical cycles with the potential for feeding back on radiative forcing (Gehlen et al., 2011; Hopkins et al., 2020). The likelihood for major disruptive impacts of ocean acidification on marine ecosystems, if future  
40  $\text{CO}_2$  emissions were to go unabated, is reflected by the Sustainable Development Goal 14.3 (SDG 14.3) - "Reduce Ocean Acidification: minimize and address impacts of ocean acidification" (<https://www.globalgoals.org/14-life-below-water>, last access: 20/03/2023). Albeit not specifically mentioned, moving towards SDG 14.3 implies the understanding of historical and contemporary carbonate chemistry, its mean state, trends and variability.

Earth system models have been widely used to track changes in ocean  $\text{pH}$  over the historical period and to project its future evolution under different  $\text{CO}_2$  emission pathways (Bopp et al., 2013; Gattuso et al., 2015; Kwiatkowski et al., 2020; Cooley  
45 et al., 2022; Jiang et al., 2023). The present-day global surface ocean  $\text{pH}$  is roughly 0.1  $\text{pH}$  units less than at the beginning of the industrial era (Gattuso et al., 2015; Jiang et al., 2019) corresponding to an increase in hydrogen ion concentration of 26% (Doney, 2010). By the end of the 21<sup>st</sup> century, the  $\text{pH}$  is projected to decrease by  $0.16 \pm 0.002$   $\text{pH}$  units in response to the IPCC AR6 low emission scenario (SSP1-2.6), respectively by  $0.44 \pm 0.005$   $\text{pH}$  units in response to the IPCC AR6 high emission pathway (SSP5-8.5) relative to 1870–1899 (Kwiatkowski et al., 2020). Understanding impacts on marine biota requires to  
50 move towards finer spatial and temporal scales than resolved by the current generation of Earth system models (Torres et al., 2021), as well as to expand the analysis from  $\text{pH}$  to other carbonate system variables such as the saturation state with respect to calcium carbonate minerals and the buffer capacity. The development and implementation of environmental management strategies equally rely on understanding and attributing the variability of the carbonate system from diurnal to decadal time scales to underlying physical-chemical-biological processes.

55 In situ time series have played an important role in monitoring ocean acidification over the last decades (Bates et al., 2014; Lauvset et al., 2015; Sutton et al., 2019; Pérez et al., 2021; Leseurre et al., 2022; Skjelvan et al., 2022). At these sites, seawater  $\text{pH}$  has been either directly measured or calculated from measurements of other carbonate system variables. These variables



include surface ocean partial pressure of CO<sub>2</sub> ( $p\text{CO}_2$ ), total alkalinity ( $A_T$ ), and dissolved inorganic carbon (DIC). While changes in time series of carbonate system variables well reflect impacts of enhanced anthropogenic CO<sub>2</sub> uptake on ocean chemistry at a local scale (Steinberg et al., 2001; González-Dávila and Santana-Casiano, 2009; Dore et al., 2009; Bates et al., 2014; Pérez et al., 2021), the reliable upscaling to large ocean regions or entire basins requires a significant extension of the existing observing network (Lauvset et al., 2015; Bakker et al., 2016; Sutton et al., 2019; Lauvset et al., 2022).

Time series data are completed by bottle data from international cruises. These data are synthesized by the Global Ocean Data Analysis Project v2.2022 (GLODAPv2.2022) and include about 1.4 million measurements of surface-to-interior ocean  $p\text{H}$ ,  $A_T$ , DIC, and other parameters (Lauvset et al., 2022, <https://www.glodap.info/>, last access: 30/9/2022). Likewise, underway measurements of near-surface CO<sub>2</sub> fugacity, i.e.,  $p\text{CO}_2$  corrected for non-ideal gas behavior, are compiled in the Surface Ocean CO<sub>2</sub> Atlas (SOCAT) since its first release in 2011 (Pfeil et al., 2013). That latest version SOCATv2022 yields approximately 33.7 million high-quality controlled data (Bakker et al., 2022, <http://www.socat.info/>, last access: 17/6/2022). Despite millions of observations available, data coverage is still modest, e.g., CO<sub>2</sub> fugacity samples over the global ocean cover less than 2% of its surface for each month in the last three decades (Bakker et al., 2016; Hauck et al., 2020). Mapping methods have become an essential tool in ocean carbon cycle research allowing to extrapolate these sparse measurements into space-time varying fields of carbonate system variables (e.g., Rödenbeck et al., 2015) and used for global carbon budget estimates (Friedlingstein et al., 2022).

Recent years have seen the rapid development of machine learning approaches to map global surface ocean  $p\text{CO}_2$  (see Rödenbeck et al., 2013; Landschützer et al., 2016; Denvil-Sommer et al., 2019; Gregor et al., 2019; Chau et al., 2022b, for instance). Thanks to these efforts, the carbon cycle community can now draw on an ensemble of reconstructions for the observation-based assessment of the ocean carbon sink (Friedlingstein et al., 2022). However, only a few global observation-based reconstructions are available for  $p\text{H}$ ,  $A_T$ , DIC, and  $\Omega$  with respect to calcite and aragonite (see Gregor and Gruber, 2021, for a review). The reconstruction of global distributions of these variables is hampered by an insufficient amount of direct measurements (Bakker et al., 2016; Lauvset et al., 2022). Alternatively, the complete carbonate system can be obtained by speciation given the information of any couple of  $p\text{CO}_2$ ,  $p\text{H}$ ,  $A_T$  or DIC together with chemical (e.g., phosphate, silicate, nitrate) and physical variables (e.g., temperature, salinity), as well as corresponding dissociation constants (Park, 1969; Lewis and Wallace, 1998; Dickson et al., 2007).

Regardless of the developments in different observation-based estimation methods, Takahashi et al. (2014), Iida et al. (2021), and Gregor and Gruber (2021) propose global climatologies or monthly varying fields of all variables of the carbonate system, i.e.,  $p\text{CO}_2$ ,  $p\text{H}$ ,  $A_T$ , DIC, and  $\Omega$ . These data products have a spatial resolution of 1° (~100km × 100km) or even coarser. Nevertheless, the variations of carbonate system variables over the coastal regions where their instantaneous gradients are driven by smaller-scale features like ocean upwelling, wind turbulence, eddies, water runoff, and sharp biological productivity (Jones et al., 2012; Bakker et al., 2016; Laruelle et al., 2017) are poorly described at such spatial resolutions. Here we improve on existing studies by providing a global 0.25°, monthly observation-based surface ocean carbonate system product consisting of datasets of six core variables of the marine carbonate system (see Table 1 and Appendix A for definitions) and their associated 1 $\sigma$ -uncertainty. This high-resolution data product covers the years from 1985 to 2021. Laboratoire des Sciences du Climat et de



l'Environnement (LSCE) is in charge of the product within the European Copernicus Marine Environment Monitoring Service (CMEMS). Our product is referred to as CMEMS-LSCE hereafter.

95 The reconstruction of surface ocean carbonate system variables starts with the reconstruction of surface ocean  $p\text{CO}_2$  and  $A_T$  in each regular grid of  $1\text{month} \times 0.25^\circ \times 0.25^\circ$ . Next, variables  $p\text{H}$ , DIC, and  $\Omega$  are derived by speciation. Advantages of the combination of  $p\text{CO}_2$  and  $A_T$  over others for the speciation of the carbonate system are: (1)  $p\text{CO}_2$  is the most extensively measured parameter, (2)  $A_T$  can be accurately predicted from salinity, temperature, and nutrient concentrations, and (3) the combination of these two prior variables results in the slightest uncertainty of  $p\text{H}$  estimates (Zeebe and Wolf-Gladrow, 2001; 100 Lauvset and Gruber, 2014; Takahashi et al., 2014; Orr et al., 2018). The three main successive modules used in the CMEMS-LSCE production chain are summarized as follows.

**Table 1.** CMEMS-LSCE carbonate system variables.

Standard names	Notations	Units
1. Partial pressure of $\text{CO}_2$ in surface seawater	$p\text{CO}_2$	$\mu\text{atm}$
2. Total alkalinity in surface seawater	$A_T$	$\mu\text{mol kg}^{-1}$
3. Surface ocean dissolved inorganic carbon	DIC	$\mu\text{mol kg}^{-1}$
4. Surface seawater $p\text{H}$ reported on total scale	$p\text{H}$	-
5. Saturation state for surface seawater with respect to aragonite	$\Omega_{ar}$	-
6. Saturation state for surface seawater with respect to calcite	$\Omega_{ca}$	-

- 105 i) *Reconstruction of  $p\text{CO}_2$*  (Sect. 3.1): a modified version of the CMEMS-LSCE-FFNN approach (Chau et al., 2022b) is applied to map gridded datasets of SOCATv2022  $\text{CO}_2$  fugacity and predictors in order to reconstruct  $p\text{CO}_2$  at a spatial resolution of  $0.25^\circ$  for every month in the period 1985-2021 (444 months in total). The CMEMS-LSCE-FFNN works on an ensemble of 100 feed-forward neural networks (FFNNs). By design, 100-member ensemble model outputs allow to yield the best model estimate (i.e., ensemble mean) and model uncertainty (i.e., ensemble standard deviation) for surface ocean  $p\text{CO}_2$  in each grid cell and each month. The primary modification of this study's version and the original CMEMS-LSCE-FFNN (Chau et al., 2022b) is an increase of 16-fold in model spatial resolution.
- 110 ii) *Reconstruction of  $A_T$*  (Sect. 3.2): locally interpolated alkalinity regression (LIAR; Carter et al., 2016, 2018) is chosen to estimate total alkalinity on regular grids of  $1\text{month} \times 0.25^\circ \times 0.25^\circ$  over the global surface ocean for the years 1985-2021. LIAR works with multiple linear regression models, each representing a combination of predictor variables. The best linear model, which has the lowest prediction error among the others, is retained for the final estimation of





$A_T$ . Various reconstruction methods for  $A_T$  exist (see Carter et al., 2016; Broullón et al., 2019; Gregor and Gruber, 2021, for a review), but we choose LIAR due to its global applicability, simplicity in setting, and accuracy compared to other published approaches (Carter et al., 2018; Gregor and Gruber, 2021). Importantly, LIAR allows determining reconstruction uncertainty propagated from multiple sources of input uncertainties at desired model resolutions.

iii) *Reconstruction of pH, DIC, and saturation states with respect to aragonite ( $\Omega_{ar}$ ) and calcite ( $\Omega_{ca}$ )* (Sect. 3.3): CO2SYS (Lewis and Wallace, 1998; Van Heuven et al., 2011) is a standard software used for the speciation of carbonate parameters in the marine  $CO_2$  system (see Olsen et al., 2016; Bresnahan et al., 2021; Gregor and Gruber, 2021; Woosley, 2021, for a few). The CO2SYS speciation is built on a set of equilibrium equations (Dickson et al., 2007; Dickson, 2010). Given the reconstructed  $pCO_2$  and  $A_T$ , non- $CO_2$  acid-base constituents, physical variables, and equilibrium constants, this method allows solving pH, DIC,  $\Omega_{ar}$ , and  $\Omega_{ca}$  at the same input resolutions. A complementary of the CO2SYS software developed by Orr et al. (2018) is used to quantify the uncertainty associated with these carbonate system variables. All the input data uncertainties are propagated through the CO2SYS processing chain.

The global monthly,  $0.25^\circ$ -resolution datasets of  $pCO_2$ ,  $A_T$ , pH, DIC,  $\Omega_{ar}$ , and  $\Omega_{ca}$  are intensively evaluated against different observation-based products independent from our model fitting at a global scale to in situ locations (Table 3). In Section 4, multiple metrics are proposed for product analyzes and assessments. Results are presented in section 5 with emphasis on the evaluation of the best reconstruction and associated model uncertainty for each variable (Sect. 5). This section also highlights the advantages obtained with an increase in spatial resolution and presents an application of the CMEMS-LSCE product in tracking ocean acidification over the last three decades. Section 6 summarizes key results, discusses the potential for future model upgrades, and introduces possible product use cases.

## 2 Data used and reprocessing

### 2.1 Input data products for surface ocean carbonate system reconstructions

Many observation-based products are used as predictors of our target carbonate system variables (Table 2). Global ocean maps of sea surface temperature (SST), salinity (SSS), height (SSH), chlorophyll-*a* (Chl-*a*) come from the Copernicus Marine Environment Monitoring Service (CMEMS: Good et al., 2020; Nardelli et al., 2016; Droghei et al., 2018; Maritorena et al., 2010). Mixed layer depth (MLD) fields belong to Estimating the Circulation and Climate of the Ocean project Phase II (ECCO2, Menemenlis et al., 2008).  $CO_2$  mole fractions ( $xCO_2$ ) are derived from the  $CO_2$  atmospheric inversion of the Copernicus Atmosphere Monitoring Service (CAMS, Chevallier et al., 2005, 2010; Chevallier, 2013). Surface ocean concentrations of nitrate ( $NO_3$ ), silicate ( $SiO_2$ ), and phosphate ( $PO_4$ ) are extracted from the World Ocean Atlas 2018 (WOA18, Garcia et al., 2019). The climatological  $pCO_2$  ( $pCO_2^{clim}$ ) product is provided by Lamont Doherty Earth Observatory (LDEO, Takahashi et al., 2009). Details of these products including resource access, data coverage, and resolutions are presented in Table 2.

With the exception of  $xCO_2$ , nutrient concentrations, and  $pCO_2^{clim}$ , these input data products have original resolutions equivalent to or even finer than a spatial resolution of  $0.25^\circ$  and a temporal resolution of monthly. When mismatches in data



**Table 2.** Input data used in the reconstructions of CMEMS-LSCE carbonate system variables over the global ocean in 1985-2021.

Variables	Notations	Units	Products	Resolutions	References
1. CO <sub>2</sub> fugacity	$f\text{CO}_2$	$\mu\text{atm}$	Surface Ocean CO <sub>2</sub> Atlas version 2022 (SOCATv2022, 1985-2021)	monthly, 1° (open ocean) and 0.25° (coastal ocean)	Bakker et al. (2022)
2. Sea surface temperature	SST	°C	CMEMS SST_GLO_SST_L4_REP_OBSERVATIONS_010_011 and SST_GLO_SST_L4_NRT_OBSERVATIONS_010_001 (1985-2021)	daily, 0.05°	Good et al. (2020)
3. Sea surface salinity	SSS	PSU	CMEMS MULTIOBS_GLO_PHY_S_SURFACE_MYNRT_015_013 (1993-2021)	monthly, 0.25°	Nardelli et al. (2016); Droghei et al. (2018)
4. Sea surface height	SSH	m	CMEMS SEALEVEL_GLO_PHY_L4_MY_008_047 and SEALEVEL_GLO_PHY_L4_NRT_OBSERVATIONS_008_046 (1993-2021)	daily, 0.25°	CLS-TOULOUSE
5. Mixed layer depth	MLD	m	Estimating the Circulation and Climate of the Ocean project Phase II (ECCO2, 1992-2021)	daily, 0.25°	Menemenlis et al. (2008)
6. Chlorophyll- <i>a</i>	CHL- <i>a</i>	$\text{mg m}^{-3}$	CMEMS OCEANCOLOUR_GLO_CHL_L4_REP_OBSERVATIONS_009_082 and OCEANCOLOUR_GLO_CHL_L4_NRT_OBSERVATIONS_009_033 (1998-2021)	daily, 0.25°	GLOCOLOUR, Maritorena et al. (2010)
7. CO <sub>2</sub> mole fraction	$x\text{CO}_2$	ppm	CO <sub>2</sub> atmospheric inversion from the Copernicus Atmosphere Monitoring Service (CAMS, 1985-2021)	3-hourly, 1.9° × 3.75°	Chevallier et al. (2005, 2010); Chevallier (2013)
8. $p\text{CO}_2$ climatology	$p\text{CO}_2^{\text{lim}}$	$\mu\text{atm}$	Lamont Doherty Earth Observatory (LDEO, climatology)	monthly, 4° × 5°	Takahashi et al. (2009)
9. Nitrate 10. Silicate 11. Phosphate	NO <sub>3</sub> SiO <sub>2</sub> PO <sub>4</sub>	$\mu\text{mol kg}^{-1}$	World Ocean Atlas 2018 (WOA18, climatologies)	monthly, 1°	Garcia et al. (2019)

\* Data products 1-8 are used in the  $p\text{CO}_2$  reconstruction. Products 2-3 and 9-11 are used to compute  $A_T$ , DIC,  $p\text{H}$ ,  $\Omega_{ar}$ , and  $\Omega_{ca}$ .

\*\* Last access was on 15/4/2022 for all input databases except for SOCATv2022 data (17/6/2022) and WOA18 data (30/7/2022).

145 resolutions appear, input data products are interpolated to fit the pre-defined model resolutions. The datasets of SST and  $x\text{CO}_2$   
 - the two key variables driving global  $p\text{CO}_2$  changes (Bates et al., 2014; Gruber et al., 2019; Landschützer et al., 2019; Chau  
 et al., 2022b; Friedlingstein et al., 2022) - cover the full learning period and the whole globe as expected. The other predictor  
 data are not available before the 1990s, when new types of satellite measurements started, and one of them (i.e., Chl-*a*) does  
 not cover the high latitudes of the winter hemisphere. We therefore gap-fill the time series in an ad hoc manner, as in previous  
 150 studies (Landschützer et al., 2016; Gregor et al., 2019; Chau et al., 2022b). Monthly climatologies of SSS, Chl-*a*, and MLD  
 computed on the available data are used for each missing year. Likewise, climatologies plus linear trends of SSH following  
 global warming effects serve for the pre-1993 period. Missing Chl-*a* data in the high latitudes of the winter hemisphere are  
 replaced by the minimum concentration of Chl-*a* over the available data for the same grid cell ( $\sim 0.01 \text{ mg m}^{-3}$ ). WOA18  
 nutrients and LDEO  $p\text{CO}_2^{\text{lim}}$  are already climatologies per se and we apply them for all the analysis years 1985-2021.

155 CO<sub>2</sub> fugacity from Surface Ocean CO<sub>2</sub> ATLAS version 2022 (SOCATv2022, Bakker et al., 2022) is used as the target data  
 in our monthly  $p\text{CO}_2$  reconstructions. The SOCAT project collects and qualifies underway observations via international ves-  
 sels, moorings, or autonomous platforms. It grids the observations at spatial resolutions of 1° or 0.25° resulting in the two



major SOCAT gridded data products. The temporal resolution of these two products is monthly. While the  $1^\circ$ -data product (SOCATv2022r100) covers the global ocean, the  $0.25^\circ$  covers solely the coastal regions. The SOCAT coastal areas is within 400 km from the shoreline (Sabine et al., 2013; Bakker et al., 2016); see Fig. A1a for an illustration. To merge the two resolutions, we first duplicate the  $1^\circ$ -open-ocean SOCATv2022 data ( $\sim 2 \times 10^5$  data points) over its sixteen  $0.25^\circ$  sub-cells. This  $0.25^\circ$ -open-ocean data are then combined with the  $0.25^\circ$ -coastal-ocean SOCATv2022 data ( $\sim 4 \times 10^5$  data points) to generate a global monthly  $0.25^\circ$  ocean data product fed to our reconstruction model of  $p\text{CO}_2$  (Sect. 3.1). The merged SOCATv2022 product at monthly,  $0.25^\circ$  resolutions is referred to as SOCATv2022r025 hereafter. The assumption of open-ocean data homogeneity of  $p\text{CO}_2$  within  $1^\circ$ -grid boxes ( $\sim 100 \text{ km} \times 100 \text{ km}$ ) does not degrade the reconstruction skill over the global open ocean (see Sect. 5 for results) where  $p\text{CO}_2$  observations are spatially auto-correlated within a global median distance of  $400 \pm 250 \text{ km}$  (Jones et al., 2012). The data distribution of SOCATv2022  $\text{CO}_2$  fugacity before and after combining is shown in Fig. A2 and Table 4.

## 2.2 Product qualification and comparison

The monthly,  $0.25^\circ$ -resolution reconstructions of carbonate system variables are qualified with gridded observation-based datasets and in-situ time series which are not used in our model fitting (Table 3).

- The SOCAT data in each reconstruction month are excluded from the model fitting, which avoids overfitting and ensures fairness in the model evaluation (Chau et al., 2022b). The global monthly CMEMS-LSCE-FFNN  $p\text{CO}_2$  fields at a spatial resolution of  $0.25^\circ$  can therefore be evaluated against the  $p\text{CO}_2$  data converted from SOCATv2022  $\text{CO}_2$  fugacity (Eq. A2) at the same resolution. Doing this, CMEMS-LSCE-FFNN  $p\text{CO}_2$  is assessed with more than  $32 \times 10^5$  open-ocean data and  $4 \times 10^5$  coastal-ocean data (Table 4). The SOCATv2022 measurements have low random uncertainty (2-5  $\mu\text{atm}$ ) but the spatio-temporal sampling bias from the month and grid centers is significant (Bakker et al., 2016). The  $0.25^\circ$ -data reconstruction is also compared to its previous version with a spatial resolution of  $1^\circ$  (Chau et al., 2022a, b).
- The monthly,  $0.25^\circ$  reconstructions of  $A_T$ , DIC, and  $p\text{H}$  are qualified based on Global Ocean Data Analysis Project bottle data version 2.2022 (GLODAPv2.2022, Lauvset et al., 2022). GLODAP provides non-gridded datasets of ocean carbon variables which have been compiled and bias-corrected from water samples taken at various depths. The measurement uncertainty is  $4 \mu\text{mol kg}^{-1}$  in  $A_T$  and DIC and between 0.01 – 0.02 in  $p\text{H}$ . Only direct measurements at depths shallower than 10 m and with a flag of 2 (best quality control) are selected for this evaluation. Measurements in each box of  $1\text{month} \times 10\text{m} \times 0.25^\circ \times 0.25^\circ$  are averaged to obtain representative data of surface  $A_T$ , DIC, and  $p\text{H}$  at  $0.25^\circ$ -grid cells for months in the period 1985-2021. This results in roughly  $16 \times 10^3$  data points for  $A_T$  and DIC over the global ocean (Table 5). Only half of that amount stems from direct  $p\text{H}$  measurements. Another half, referred to as indirect measurements (i.e.,  $p\text{H}$  calculated with  $A_T$  and DIC), is excluded from this data evaluation. Over 30% of these data are distributed along the coasts. The number of the GLODAPv2.2022 gridded data (Table 5) is much less than the SOCATv2022 gridded data (Table 4).



190 • In situ time series of direct measurements of carbonate system variables ( $p\text{CO}_2$ ,  $A_T$ , DIC, and  $p\text{H}$ ) are used to qualify our  
product at local scale (Table 3). Sutton et al. (2019) present data over multiple sites equipped with autonomous moorings  
measuring surface ocean  $p\text{CO}_2$  and  $p\text{H}$  from the open ocean to the continental shelves since 2004. These time series  
were used to qualify the CMEMS-LSCE-FFNN reconstruction in Chau et al. (2022b). This study only revisits eight  
195 coastal sites (Table A2) where both  $p\text{CO}_2$  and  $p\text{H}$  have been measured and the  $1^\circ$ -reconstruction poorly constrains most  
of these measurements (see later in Sect. 3.1). The eight stations are located along the US coast, the Gulf of Mexico,  
and in a Caribbean coral reef (Fig. A1b and Table A2). Measurement uncertainty is up to  $2 \mu\text{atm}$  reported for  $p\text{CO}_2$   
and 0.02 for  $p\text{H}$ . For  $A_T$  and DIC, we consider four time series: (1) Bermuda Atlantic Time Series (BATS, Michaels and  
Knap, 1996; Steinberg et al., 2001), (2) Atmospheric Flux Dynamics Time Series in the Mediterranean (DYFAMED,  
Coppola et al., 2021), (3) European Station for Time-Series in the Ocean Canary islands (ESTOC, González-Dávila and  
200 Santana-Casiano, 2009), and (4) Hawaii Ocean Time-series (HOT, Dore et al., 2009). The first three stations are in the  
North Atlantic while the latter is located in the North Pacific (Fig. A1b). These long-term time series provide insights  
into changes in the surface ocean carbonate system over the recent decades (Bates et al., 2014; Coppola et al., 2020;  
Gregor and Gruber, 2021; Pérez et al., 2021). The HOT and ESTOC stations provide surface ocean observations of  $A_T$   
and DIC, and BATS and DYFAMED data are extracted at seawater depth shallower than 10 m. A monthly average is  
205 applied for all the mentioned time series in order to be compatible with output from the CMEMS-LSCE chain of models.

### 3 Reconstruction methods

#### 3.1 Ensemble $p\text{CO}_2$ mapping feed-forward neural networks

The CMEMS-LSCE-FFNN (Chau et al., 2022b) is based on an ensemble of 100 feed-forward neural network (FFNN) models  
mapping SOCAT  $\text{CO}_2$  fugacity ( $f\text{CO}_2$ ) and predictor variables (Eq. 1).

$$210 \quad f\text{CO}_2 = \text{FFNN} (\text{SST}, \text{SSS}, \text{SSH}, \text{Chl} - a, \text{MLD}, x\text{CO}_2, f\text{CO}_2^{\text{clim}}, \text{latitude}, \text{longitude}) \quad (1)$$

The predictors of  $f\text{CO}_2$  include sea surface temperature (SST), salinity (SSS), surface height (SSH), chlorophyll- $a$  (Chl- $a$ ),  
mixed layer depth (MLD),  $\text{CO}_2$  mole fraction ( $x\text{CO}_2$ ),  $f\text{CO}_2$  climatologies ( $f\text{CO}_2^{\text{clim}}$ ), and the geographical coordinates  
(latitude and longitude). The datasets of SOCAT  $f\text{CO}_2$  and predictors are first reprocessed to match model fitting requirements  
(Sect. 2.1). After excluding the data in the reconstruction month, the data within the 3-month window are separated into FFNN  
215 training and validation subsets with a ratio of 2 : 1. The excluded SOCATv2022 datasets are used in model evaluation. The  
CMEMS-LSCE-FFNN approach was originally developed for  $p\text{CO}_2$  reconstructions at monthly,  $1^\circ$  resolutions where  $p\text{CO}_2$   
is converted from  $f\text{CO}_2$  following the formulation (A2) by Körtzinger (1999). The model best estimate and its uncertainty are  
defined as the ensemble mean ( $\mu$ ) and ensemble spread ( $\sigma$ ) of 100 model outputs of  $p\text{CO}_2$ .

This study slightly modifies the CMEMS-LSCE-FFNN ensemble approach by Chau et al. (2022b) to achieve  $p\text{CO}_2$  recon-  
220 structions at monthly,  $0.25^\circ$  resolutions. Some of the input datasets presented here (Table 2) are different from those presented



**Table 3.** Data sources used in product evaluation and comparison.

	Product	Data type	Evaluation variables	Reference
Global ocean	1. Surface Ocean CO <sub>2</sub> Atlas version 2022 (SOCATv2022, 1985-2021), last access: 17/6/2022	observation-based gridding, resolution: 1° (global ocean) and 0.25° (coastal ocean), monthly	<i>p</i> CO <sub>2</sub>	Bakker et al. (2022)
	2. CMEMS global ocean surface carbon product (MULTI-OBS_GLO_BIO_CARBON_SURFACE_REP_015_008, 1985-2021), last access: 05/12/2022	SOCAT-based reconstruction, resolution: 1°, monthly	<i>p</i> CO <sub>2</sub>	Chau et al. (2022a, b)
	3. Global Ocean Data Analysis Project bottle data version 2.2022 (GLODAPv2.2022, 1985-2021), last access: 30/9/2022	observation	A <sub>T</sub> , DIC, pH	Lauvset et al. (2022)
Time series stations	4. Autonomous time series from surface buoys since 2004 (see details in Table A2), last access: 15/10/2022	observation	<i>p</i> CO <sub>2</sub>	Sutton et al. (2019)
	5. Bermuda Atlantic Time Series (BATS, 31.7°N-64.2°W, 1988-2021), last access: 30/10/2022	observation	A <sub>T</sub> , DIC	Michaels and Knap (1996); Steinberg et al. (2001)
	6. Atmospheric Flux Dynamics Time Series in the Mediterranean (DY-FAMED, 43.5°N-7.9°E, 1998-2017), last access: 23/03/2023	observation	A <sub>T</sub> , DIC	Coppola et al. (2020, 2021)
	7. European Station for Time-Series in the Ocean Canary islands (ES-TOC, 29.2°N-15.5°W, 1995-2009), last access: 30/10/2022	observation	A <sub>T</sub> , DIC	González-Dávila and Santana-Casiano (2009)
	8. Hawaii Ocean Time-series (HOT, 22.5°N-158.1°W, 1988-2020), last access: 30/10/2022	observation	A <sub>T</sub> , DIC	Dore et al. (2009)

in Chau et al. (2022b) (Table S1). The up-to-date input datasets have higher resolutions and a better coverage over the coastal ocean as well as in the high latitudes. Furthermore, the new CMEMS data resources offer space-time varying uncertainty fields which are important in quantifying carbonate system variable uncertainties.

For comparable evaluations in this study, we execute 100-member ensembles of FFNN models at spatial resolutions of both 1° (FFNNr100) and 0.25° (FFNNr025) using the same lot of input data resources (Table 2). Remind that the training data of *f*CO<sub>2</sub> is extracted from the SOCATv2022r100 product for FFNNr100 while it comes from the SOCATv2022r025 product (i.e., the merged product of the 1°-open-ocean dataset and the 0.25°-coastal-ocean dataset) for FFNNr025. All input datasets are reprocessed with respect to each model resolution (Sect. 2.1). Sect. 3.1 compares these two CMEMS-LSCE-FFNN versions and highlights the skill of the finer resolution data product.

### 230 3.2 Locally interpolated alkalinity regression

Locally interpolated alkalinity regression (LIAR; Carter et al., 2016, 2018) is an ensemble-based regression method developed for the global reconstruction of total alkalinity (A<sub>T</sub>). Regression coefficients were learned on GLODAPv2 data (Olsen et al., 2016) binned within regular windows of 5° × 5°. For prediction, the LIAR software interpolates between the regression



coefficients to arbitrary resolutions specified by the users. This study employs eight LIAR models (Carter et al., 2018, Table 2)  
235 for calculating  $A_T$  at monthly,  $0.25^\circ$  resolutions. Each model represents a combination of predictor variables (see the full  
presentation in Eq. 2),

$$A_T = \text{LIAR}(\text{SSS}, \text{SST}, \text{NO}_3, \text{SiO}_2). \quad (2)$$

The eight regression models include salinity (SSS) - the predominant predictor of  $A_T$  - and some combinations of temperature  
(SST), nitrate ( $\text{NO}_3$ ), and silicate ( $\text{SiO}_2$ ). The model which has the smallest propagation uncertainty is chosen to provide the  
240 best estimate of  $A_T$ .

Global monthly total alkalinity and  $1\sigma$ -uncertainty are estimated with given input data from the monthly CMEMS SSS  
and SST fields and from the WOA18 datasets of nutrient concentrations (Table 2). Uncertainty of the  $A_T$  field is estimated  
systematically through input uncertainty propagation along the processing chain (Carter et al., 2018). Here we define the input  
uncertainty of predictors in terms of standard deviations ( $1\sigma$ ). Input uncertainty fields associated to the monthly CMEMS  
245 SSS and SST are products' analysis errors (see e.g., Fig. A8) while uncertainties of the WOA18  $\text{NO}_3$  and  $\text{SiO}_2$  climatologies  
are set to 15% of data values per cell. The 15% quantity refers to the median percentage of standard analysis errors against  
climatological means of nutrient concentrations (see product standard errors in Table 6, Garcia et al., 2019). The WOA18  
standard analysis errors are defined as misfits between their interpolated data and GLODAPv2 bottle data (Olsen et al., 2016).  
Spatial distribution of the error percentage of the WOA18 nutrient concentrations at the ocean surface is illustrated in Fig. A7.

### 250 3.3 Carbonate system speciation

The CO2SYS speciation software was first developed by Lewis and Wallace (1998) to determine carbonate system parameters  
in the marine  $\text{CO}_2$  system based on a set of equilibrium equations (Dickson et al., 2007). Here we use the speciation program  
written by Van Heuven et al. (2011) and its extension with uncertainty propagation proposed by Orr et al. (2018). To obtain a  
complete description of the ocean carbonate system, the CO2SYS initialization requires the following input conditions:

- 255
- i) values of any couple of the parameters  $p\text{CO}_2$ ,  $A_T$ , DIC, and  $p\text{H}$ ,
  - ii) temperature and pressure,
  - iii) total concentrations of all the non- $\text{CO}_2$  acid-base systems,
  - iv) equilibrium constants used to describe seawater acid-base chemistry.

The (iii)-condition involves total concentrations of both conservative and non-conservative constituents in the non- $\text{CO}_2$  acid-  
260 base systems. The amount of conservative constituents such as borate, fluoride, and sulfate in surface seawater is estimated with  
salinity. The total concentration of non-conservative constituents (nutrients) is computed approximately with silicate ( $\text{SiO}_2$ ),  
and phosphate ( $\text{PO}_4$ ). Further information of the carbonate system speciation can be found in Dickson et al. (2007) and Dickson  
(2010).





With the reconstructions of  $p\text{CO}_2$  and  $A_T$  (Sects 3.1 and 3.2), the CO2SYS speciation software is used to derive  $p\text{H}$ , DIC,  $\Omega_{ar}$ , and  $\Omega_{ca}$ , and determine their uncertainty over the ocean surface at a resolution of  $0.25^\circ$ . Equation 3 expresses all input-output variables of CO2SYS for this study. Note that the estimates for other carbonate system variables such as hydrogen ion ( $H^+$ ) concentration and Revelle Factor (RF) - a measure of the carbonate buffer capacity- are also available (Figs. A4 and A6) but beyond the scope of our data evaluation.

$$p\text{H}, \text{DIC}, \Omega_{ar}, \Omega_{ca} = \text{CO2SYS}(p\text{CO}_2, A_T, \text{SST}, \text{SSS}, P, \text{SiO}_2, \text{PO}_4, \text{constants}) \quad (3)$$

The FFNN best estimate (ensemble mean) of  $p\text{CO}_2$  reconstructions (Sect. 3.1) and the LIAR outputs of  $A_T$  (Sect. 3.2) are used as the prior inputs of the CO2SYS at each grid cell for every month in the period 1985-2021. We take the same data products of SST, SSS, and nutrient concentrations as for the previous reconstructions (Table 2). Pressure (P) is assumed to be 0 dbar at the ocean surface. For equilibrium constants, we choose the best empirical values recommended by Dickson et al. (2007) and Dickson (2010). These settings include (1) the dissociation constants  $K_1$  and  $K_2$  from Lueker et al. (2000) and  $K_{\text{HSO}_4}$  from Dickson (1990) in combination with the total boron-ratio-salinity formulation by Uppstrom (1974).

The uncertainty of the CO2SYS variables is estimated by error propagation (Orr et al., 2018). Inputs for the CO2SYS error propagation include the reconstruction uncertainty of  $p\text{CO}_2$  (FFNN ensemble standard deviation) and of  $A_T$  (LIAR error propagation). The uncertainty of SST, SSS, and nutrient concentrations are set to the same values as in the previous section (Sect. 3.2). Equilibrium constants' standard errors are default values (see Table 1, Orr et al., 2018). As for FFNN and LIAR, uncertainty values of each carbonate system variable are computed for each month in 1985-2021 and at each  $0.25^\circ$ -grid box over the global surface ocean.

## 4 Evaluation metrics

### 4.1 Model best estimate and uncertainty quantification

The 100 FFNN models result in an ensemble of 100 estimates of global monthly,  $0.25^\circ$  surface ocean  $p\text{CO}_2$  fields (Sect. 3.1). Specify any  $t = 1 : 444$  (month),  $i = 1 : 180$  (latitude), and  $j = 1 : 360$  (longitude), the best estimate ( $\mu_{tij}$ ) and uncertainty ( $\sigma_{tij}$ ) at time  $t$  and grid cell  $ij$  are deduced from 100 FFNN  $p\text{CO}_2$  estimates  $(X(t, i, j, m))_{m=1}^{m=100}$  as follows.

$$\mu_{tij} = \frac{\sum_{m=1}^{m=100} X(t, i, j, m)}{100}, \quad (4a)$$

$$\sigma_{tij} = \sqrt{\frac{\sum_{m=1}^{m=100} [X(t, i, j, m) - \mu_{tij}]^2}{100}}. \quad (4b)$$

For  $p\text{H}$ ,  $A_T$ , DIC,  $\Omega_{ar}$ , and  $\Omega_{ca}$ , the best estimates and associated uncertainties ( $\mu_{tij}$  and  $\sigma_{tij}$ ) are obtained directly from the LIAR and CO2SYS speciation tools and their error propagation (Sects. 3.2 and 3.3).

To assign representatives of  $\mu$  and  $\sigma$  estimates for carbonate system variables at a specific space-time window, we define statistics with respect to each of the three following cases:



i) a representative over a period of time ( $T$  months)

$$\mu_{ij} = \frac{\sum_t \mu_{tij}}{T}, \quad (5a)$$

$$\sigma_{ij} = \sqrt{\frac{\sum_t \sigma_{tij}^2}{T}}. \quad (5b)$$

ii) a representative over a region (e.g., ocean basins and sub-basins, the global ocean)

$$\mu_t = \frac{\sum_{ij} \mu_{tij} \times A_{ij}}{\sum_{ij} A_{ij}}, \quad (6a)$$

$$\sigma_t = \sqrt{\frac{\sum_{ij} \sigma_{tij}^2 \times A_{ij}}{\sum_{ij} A_{ij}}}. \quad (6b)$$

iii) a representative over a period of time and a region

$$\mu = \frac{\sum_{t,ij} \mu_{tij} \times A_{ij}}{T \times \sum_{ij} A_{ij}}, \quad (7a)$$

$$\sigma = \sqrt{\frac{\sum_{t,ij} \sigma_{tij}^2 \times A_{ij}}{T \times \sum_{ij} A_{ij}}}. \quad (7b)$$

where  $t$  is in a time period with length  $T$  and  $A_{ij}$  is the area of each grid cell in a desired region. It is noteworthy that the statistics in Eqs. (5b)-(7b) are not the standard deviation associated to the mean quantities in Eqs. (5a)-(7a), but they stand for the best representative of uncertainty estimates over an ocean basin and/or time period. These statistics also support for the comparison with model-observation deviation (e.g., Eq. 10) which is typically used in the calculation of standard uncertainty proposed in the previous studies (Jiang et al., 2019; Iida et al., 2021; Gregor and Gruber, 2021). Subscripts in the notations of the best model estimates ( $\mu$ ) and model uncertainties ( $\sigma$ ) in Eqs. (4)-(7) are dropped out for general situations.

The best model estimate ( $\mu$ ) is assessed against model uncertainty ( $\sigma$ ) through  $\sigma$ -to- $\mu$  ratio (%)

$$R(\mu, \sigma) = 100\% \frac{\sigma}{|\mu|}. \quad (8)$$

The  $\sigma$ -to- $\mu$  ratio allows evaluating the significance of the model estimate. Model estimates of carbonate variables are reliable with  $R(\sigma, \mu)$  values less than 20% (Rose, 2013).

## 4.2 Model performance in comparison with evaluation data

Assume that  $\mu_{tij}$  and  $O_{tij}$  are the best model estimate and an observation (or its gridded data) available at time  $t$  and grid cell  $ij$ , and  $\mu$  and  $O$  are respectively their means over the total number of evaluation data ( $N$ ). Model skills are assessed against observation data (Table 3) with the following metrics:

- mean model-observation differences (Bias)

$$\text{Bias} = \frac{\sum_{t,ij} (\mu_{tij} - O_{tij})}{N}, \quad (9)$$



- root-of-mean-square-deviation (RMSD)

$$\text{RMSD} = \sqrt{\frac{\sum_{t,ij} (\mu_{tij} - O_{tij})^2}{N}}, \quad (10)$$

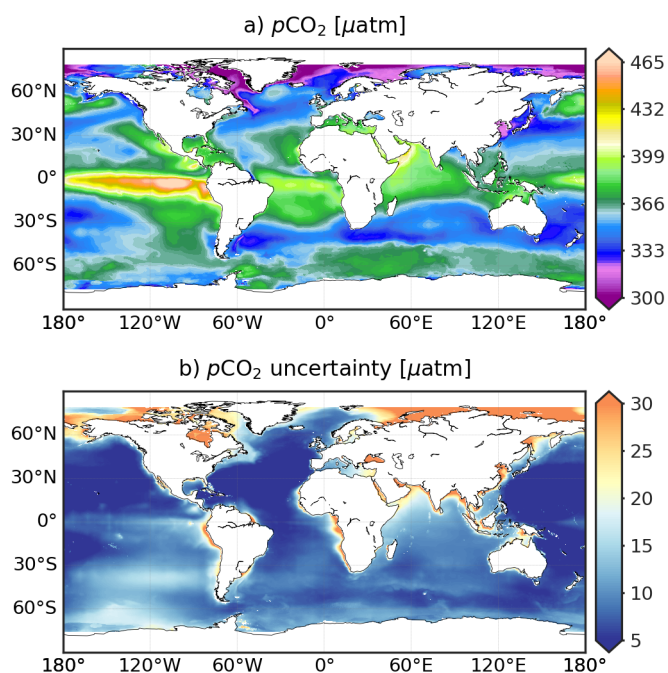
- 320
- coefficient of determination ( $r^2$ )

$$r^2 = \frac{\left[ \sum_{t,ij} (\mu_{tij} - \mu) \times (O_{tij} - O) \right]^2}{\sum_{t,ij} (\mu_{tij} - \mu)^2 \times \sum_{t,ij} (O_{tij} - O)^2}. \quad (11)$$

## 5 Results

### 5.1 Surface ocean $p\text{CO}_2$

This section presents the reconstruction of surface ocean  $p\text{CO}_2$  at monthly and  $0.25^\circ$  resolutions. The reconstruction skill is evaluated against SOCATv2022 data from global to ocean basin scale and at the level of grid cells (Sect. 2.2). We compare the novel reconstruction at a higher spatial resolution to the one at a coarser spatial resolution (Chau et al., 2022b). Emphasis is put on the skill to reproduce spatial and temporal variations of  $p\text{CO}_2$  across a variety of coastal regions and time series stations.



**Figure 1.** CMEMS-LSCE-FFNN  $p\text{CO}_2$  over the global ocean at a spatial resolution of  $0.25^\circ$ . Temporal means of the model best estimate and  $1\sigma$ -uncertainty per grid cell over 1985-2021 are calculated by using Eq. (5).



**Table 4.** Skill scores for monthly CMEMS-LSCE-FFNN reconstructions of  $p\text{CO}_2$  at  $1^\circ$  (r100) and  $0.25^\circ$  (r025) spatial resolutions computed over the period 1985-2021. r100  $\rightarrow$  025 and r025  $\rightarrow$  100 are referred to the versions upscaled or downscaled from the original CMEMS-LSCE-FFNN  $p\text{CO}_2$  at  $1^\circ$  and  $0.25^\circ$  resolutions. SOCATv2022 gridded data independent from CMEMS-LSCE-FFNN training are used as benchmarks for model evaluation (see text for details). Statistics including total numbers of data, RMSD (Eq. 10), and  $r^2$  (Eq. 11) are reported for both the open ocean (O) and coastal region (C). \* marks results with respect to the primary product proposed in this study.

Basins			Number of data		RMSD [ $\mu\text{atm}$ ]				$r^2$			
			r100	r025*	r100	r025 $\rightarrow$ 100	r025*	r100 $\rightarrow$ 025	r100	r025 $\rightarrow$ 100	r025*	r100 $\rightarrow$ 025
0.	Globe	(O)	207174	3317273	14.32	14.08	14.29	14.38	0.83	0.83	0.83	0.83
		(C)	101007	431758	26.61	26.48	27.55	28.50	0.72	0.72	0.74	0.72
1.	Arctic	(O)	537	8589	27.93	27.43	28.04	28.06	0.69	0.69	0.67	0.67
		(C)	5897	25844	38.74	38.56	41.46	43.17	0.55	0.56	0.55	0.52
2.	Atlantic	(O)	54797	876116	13.76	13.57	13.69	13.78	0.81	0.81	0.81	0.81
		(C)	49770	227665	24.99	24.78	25.17	26.05	0.76	0.76	0.77	0.77
3.	Pacific	(O)	120604	1932981	14.59	14.30	14.54	14.67	0.85	0.85	0.85	0.85
		(C)	26847	104269	26.79	26.90	28.46	28.95	0.71	0.71	0.69	0.67
4.	Indian Ocean	(O)	4485	71719	10.34	10.17	10.26	10.34	0.88	0.88	0.88	0.88
		(C)	1522	6187	23.50	22.82	25.40	26.51	0.69	0.71	0.69	0.69
5.	Southern Ocean	(O)	26751	427868	14.42	14.29	14.52	14.43	0.69	0.69	0.69	0.69
		(C)	16971	67793	26.01	25.80	27.35	28.80	0.61	0.61	0.64	0.59

Figure 1 presents global maps at  $0.25^\circ$ -resolution of long-term averages of  $p\text{CO}_2$  and corresponding uncertainty estimates. Reconstructed  $p\text{CO}_2$  distributions reveal well documented large scale structures. Values are high over upwelling regions (e.g., Equatorial Pacific, California Boundary Current, Western Arabian Sea). Low  $p\text{CO}_2$  is associated with increased  $\text{CO}_2$  solubility in cold high latitudes seawater (e.g., Arctic), strong biological production (e.g., China Sea), or the combination of both (e.g., subpolar Northern Atlantic, Southern Ocean between  $35 - 50^\circ\text{S}$ ). Spatial structures appear coherent from small to large spatial scales, both along the coast and moving towards the open ocean (see also in Figs. 2-4). The combination of a down-scaled version of open-ocean and higher-resolution coastal SOCATv2022 data (Sect. 2.1) yields  $p\text{CO}_2$  distributions without discontinuities. The uncertainty map (Fig 1b) represents the confidence level in surface ocean  $p\text{CO}_2$  estimates (Fig 1a). Predominantly low uncertainty estimates ( $\sigma < 5 \mu\text{atm}$ ) indicate the global stability of the ensemble reconstruction. Exceptions are found in many coastal regions, open-ocean areas with sparse data coverage (e.g., Southern Pacific, Indian Ocean), or regions with substantially high or low surface ocean  $p\text{CO}_2$  (e.g., Arctic, eastern equatorial Pacific). However,  $p\text{CO}_2$  is reconstructed with a high degree of confidence over most of the global ocean with a  $\sigma$ -to- $\mu$  ratio (Eq. 8) below 10% (Fig. A9a).

Skill scores of the monthly,  $0.25^\circ$ -resolution reconstruction are presented in Table 4 (columns marked by an asterisk). The global RMSD (Eq. 10) between the best reconstruction and SOCATv2022r025  $p\text{CO}_2$  over the entire period is  $14.29 \mu\text{atm}$  for the open ocean and  $27.55 \mu\text{atm}$  for the coastal ocean. These two model errors are lower than 4% and 8% of the global mean  $p\text{CO}_2$  (Table 7). Moreover, variability present in observation-based data is reproduced by the CMEMS-LSCE-FFNN with high values of  $r^2$  (open ocean: 0.83, coast: 0.74). The reconstruction quality is similar among major ocean basins. Spatial distributions of SOCATv2022 data, bias, and RMSD are shown in Figs. A2-bd and A3-bdfh. Estimation skills are low in the



ocean basins with sparse data coverage and significant space-time variability of  $p\text{CO}_2$  (e.g., Arctic, eastern Equatorial Pacific, land-ocean continuum).

Table 4 also presents statistics for the monthly FFNN products of surface ocean  $p\text{CO}_2$  at spatial resolutions of  $0.25^\circ$  (r025) and  $1^\circ$  (r100) together with their variants (r100  $\rightarrow$  025 and r025  $\rightarrow$  100). The latter are respectively extrapolation and interpolation versions of the original r100 and r025 datasets, i.e., FFNN model outputs regridded to a finer or coarser spatial resolution. For compatibility, we compare statistics between:

- i) FFNN(r025) and FFNN(r100  $\rightarrow$  025) by using SOCATv2022r025 as evaluation data,
- ii) FFNN(r025  $\rightarrow$  100) and FFNN(r100) by using SOCATv2022r100 as evaluation data.

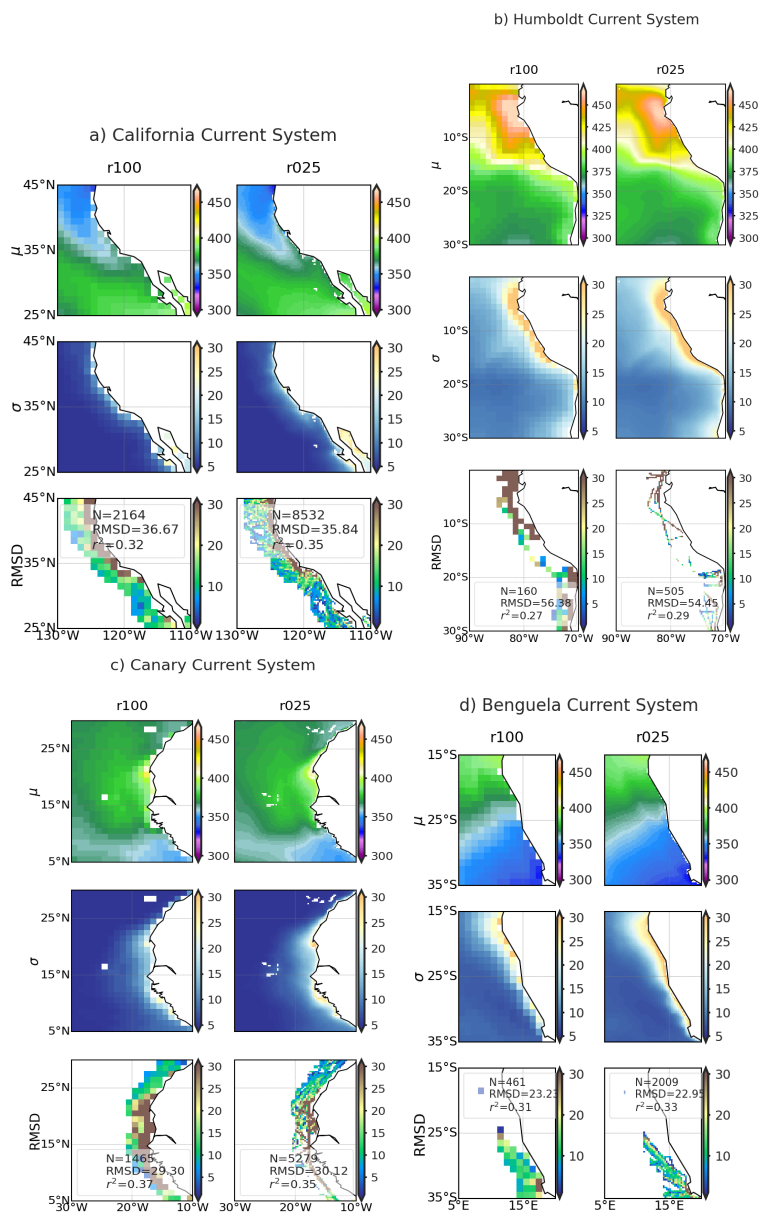
The FFNN(r025) central to this study yields a lower RMSD and a higher correlation to the SOCAT data than the FFNN(r100  $\rightarrow$  025). As expected, the improvement in reconstruction skill with higher model resolution is larger over coastal regions than in the open ocean. The FFNN(r025) product after interpolating to a coarser resolution, i.e., FFNN(r025  $\rightarrow$  100), agrees with the original  $1^\circ$ -resolution data product over all the ocean.

The motivation to increase the spatial resolution of the reconstruction is to improve the representation of horizontal gradients of  $p\text{CO}_2$  at fine scales. Figures 2-4 exemplify spatial distributions for the two reconstructions (r025 and r100) over the coastal-open-ocean continuum. Ten distinct oceanic regions are considered (see Fig. A1a and Table A1 for the ten locations), which can be classified into three groups:

- permanent Eastern Boundary current upwelling systems with relatively high  $p\text{CO}_2$  (California Current System - CCS, Humboldt Current System - HCS, Canary Current System - CnCS, and Benguela Current System - BCS),
- regions characterized by low  $p\text{CO}_2$  values driven by cold water temperatures and strong biological production (Labrador Sea, Western South Atlantic, Northern Europe, and Sea of Japan),
- other regions either under the influence of strong river runoff (Amazon mouth) or monsoon-driven upwelling (Western Arabian Sea).

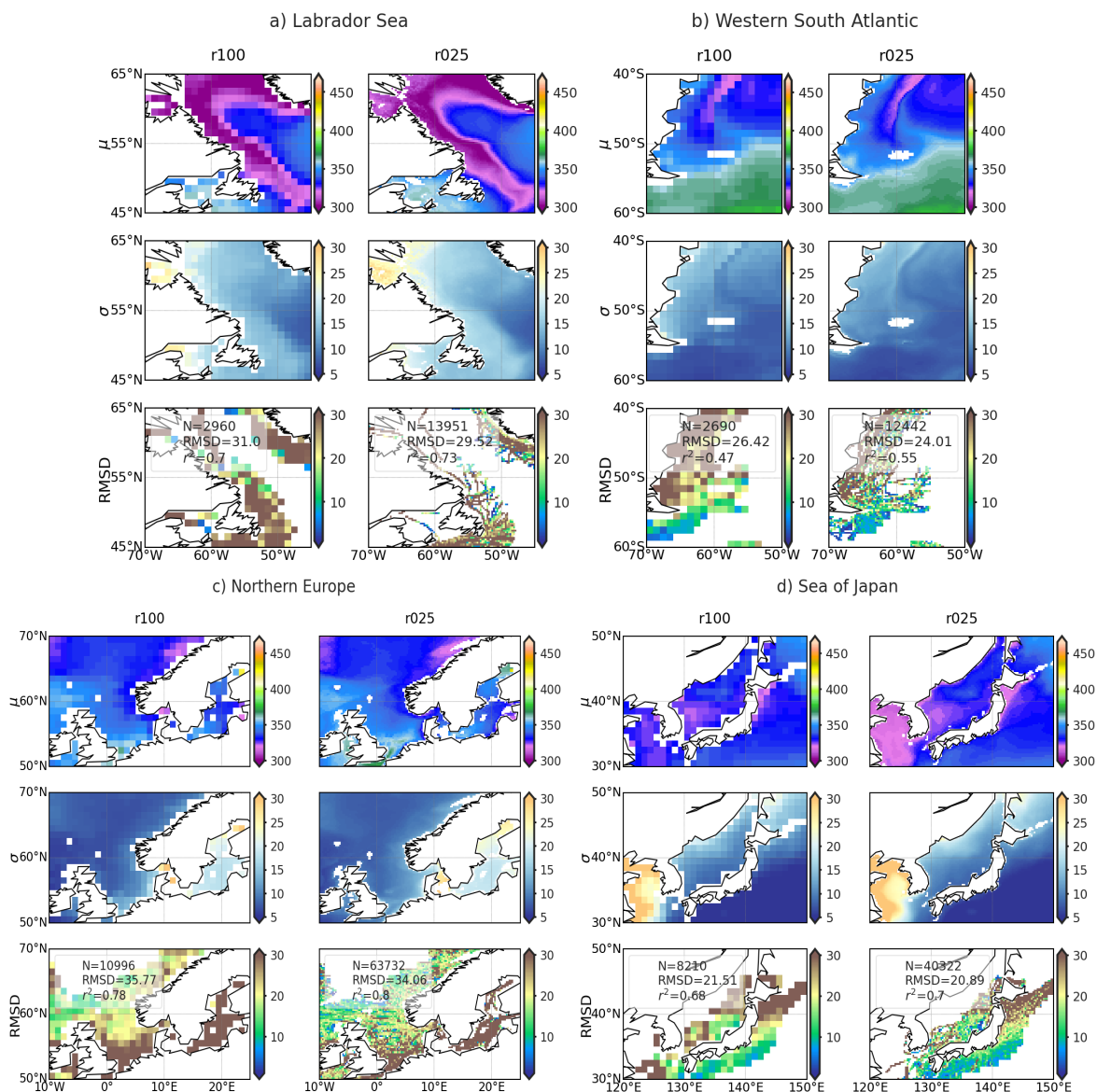
The legend of Figs 2-4 includes regional RMSD and  $r^2$  computed between the best estimates of two models and coastal-ocean SOCATv2022r025 data. The coarser spatial resolution product is co-located at the same  $0.25^\circ$ -grid cells for this analysis. These figures illustrate important discrepancies in  $p\text{CO}_2$  data density between coastal regions with poorly monitored regions (e.g., HCS, BCS, Amazon mouth) contrasting with areas with higher data coverage (e.g., Northern Europe, Sea of Japan).

Over 7 out of the 10 analysed regions the reconstruction at monthly,  $0.25^\circ$  resolutions yields RMSDs below 10% of the global mean of coastal-ocean  $p\text{CO}_2$  estimates (Table 7) and  $r^2$  values higher than 0.3; e.g., Northern Europe (RMSD =  $33.90 \mu\text{atm}$ ,  $r^2 = 0.80$ ), Sea of Japan (RMSD =  $20.84 \mu\text{atm}$ ,  $r^2 = 0.70$ ), and CnCS (RMSD =  $30.36 \mu\text{atm}$ ,  $r^2 = 0.35$ ). The CMEMS-LSCE-FFNN model projections of  $p\text{CO}_2$  lack skill over the HCS (RMSD =  $54.54 \mu\text{atm}$ ,  $r^2 = 0.29$ ), the region under influence of the Amazon river (RMSD =  $45.93 \mu\text{atm}$ ,  $r^2 = 0.37$ ), and the Western Arabian Sea (RMSD =  $45.31 \mu\text{atm}$ ,  $r^2 = 0.47$ ). In nearshore sectors of these coastal areas,  $p\text{CO}_2$  estimates are also subject to a substantial amount of uncertainty ( $\sigma > 20 \mu\text{atm}$ ).



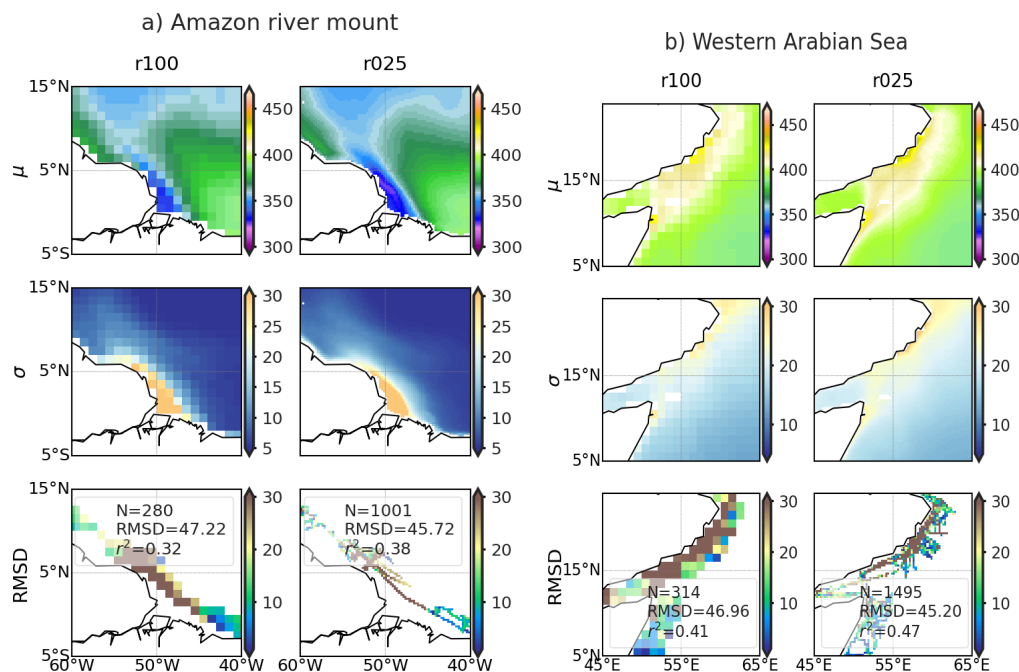
**Figure 2.** Comparison of CMEMS-LSCE-FFNN mapping  $p\text{CO}_2$  at  $1^\circ$  ( $r100$ ) and  $0.25^\circ$  ( $r025$ ) resolutions over 4 permanent upwelling regions associated with the Eastern Boundary Currents (California, Peru, Canary, and Benguela; see Figure A1-ABGH for geographical locations). For each region, spatial distributions of  $p\text{CO}_2$  ( $\mu$ ) and uncertainty ( $\sigma$ ) estimates, and coastal-ocean RMSD of  $p\text{CO}_2$  averaged over 1985-2021 (Eqs. 5 and 10) are shown. Metrics presented in the legend for each of the 3rd row include the number of coastal-ocean SOCATv2022 data ( $N$ ), regional RMSD (Eq. 10) and  $r^2$  (Eq. 11).





**Figure 3.** Comparison of CMEMS-LSCE-FFNN mapping  $p\text{CO}_2$  at  $1^\circ$  (r100) and  $0.25^\circ$  (r025) resolutions over 4 regions characterized with low  $p\text{CO}_2$  values (Labrador, South America, Northern Europe, and Japan; see Figure A1-CEFJ for geographical locations). For each region, spatial distributions of  $p\text{CO}_2$  ( $\mu$ ) and uncertainty ( $\sigma$ ) estimates, and coastal-ocean RMSD of  $p\text{CO}_2$  averaged over 1985-2021 (Eqs. 5 and 10) are showed. Metrics present in the legend for each of the 3rd row include the number of coastal-ocean SOCATv2022 data ( $N$ ), regional RMSD (Eq. 10) and  $r^2$  (Eq. 11).

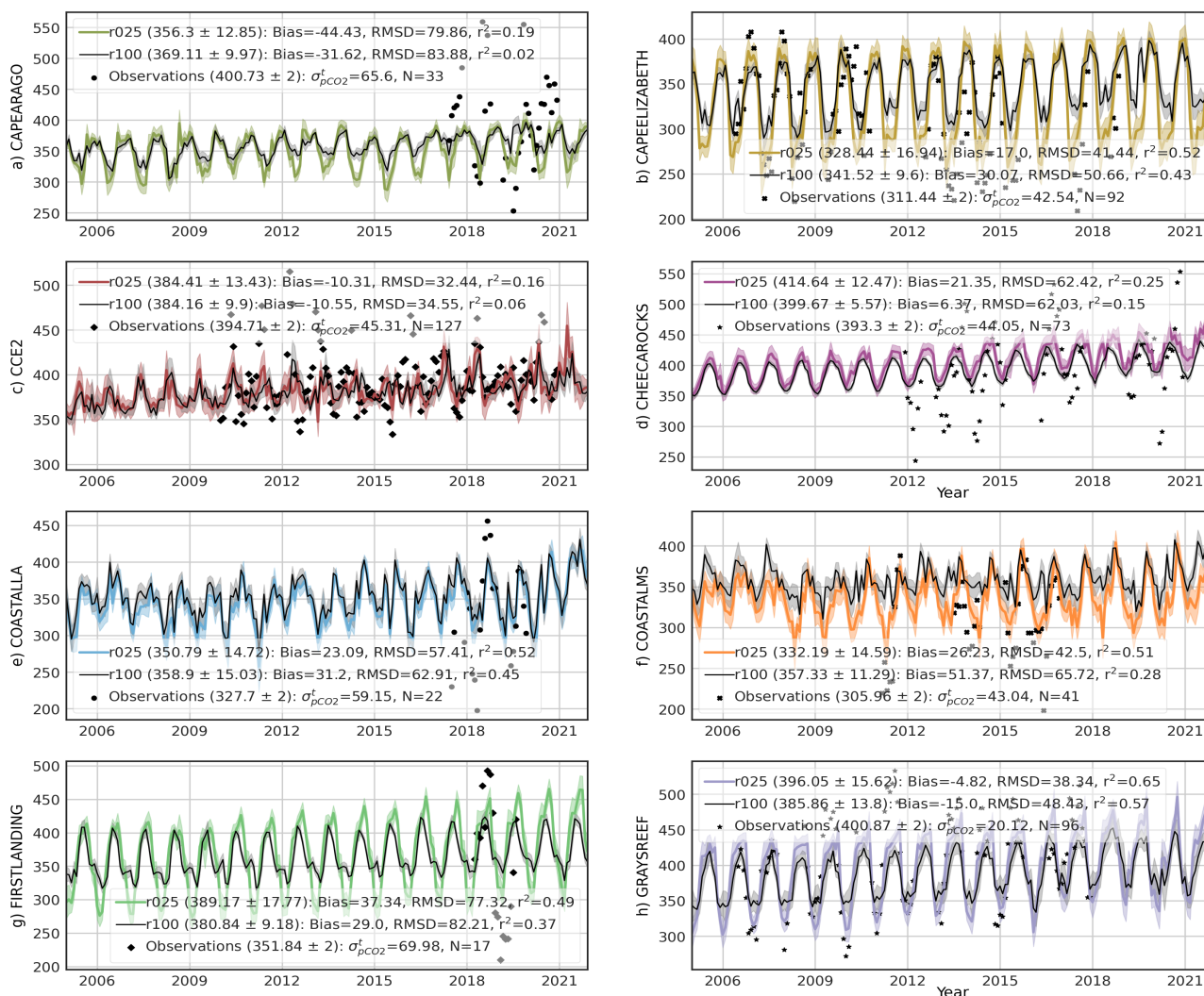
The lack of model skill reflects the combination of low data density and strong  $p\text{CO}_2$  gradients driven by multiple underlying physical and biogeochemical processes. The HCS, for instance, is characterized by the highest  $p\text{CO}_2$  levels (Fig. 2) among the



**Figure 4.** Comparison of CMEMS-LSCE-FFNN mapping  $p\text{CO}_2$  at  $1^\circ$  (r100) and  $0.25^\circ$  (r025) resolutions over the mouth of the river Amazon and the Western Arabian Sea (see Fig. A1-DI for geographical locations). For each region, spatial distributions of  $p\text{CO}_2$  ( $\mu$ ) and uncertainty ( $\sigma$ ) estimates, and coastal-ocean RMSD of  $p\text{CO}_2$  averaged over 1985–2021 (Eqs. 5 and 10) are shown. Metrics present in the legend for each of the 3rd row include the number of coastal-ocean SOCATv2022 data ( $N$ ), regional RMSD (Eq. 10) and  $r^2$  (Eq. 11).

380 four Eastern Boundary Current Systems, with interannual variability amplified with the El Niño–Southern Oscillation (ENSO) events (Feely et al., 1999; Landschützer et al., 2016). Similarly, high  $p\text{CO}_2$  levels with substantial seasonal variability are observed over the Western Arabian Sea (Fig. 4b), the key driver being monsoonal upwelling (Sabine et al., 2002; Sarma et al., 2013). In contrast to the two aforementioned coastal regions, high  $\text{CO}_2$  undersaturation as well as strong  $p\text{CO}_2$  gradients (Fig. 4a) are found in the area under the influence of Amazon river discharge (Olivier et al., 2022). Extreme values and large  
 385 variability of  $p\text{CO}_2$  challenge any approach to estimate  $p\text{CO}_2$  data over these regions (Ibáñez et al., 2015; Bakker et al., 2016; Landschützer et al., 2020).

The two FFNN reconstructions (r025 and r100) share similarities in overall structures of  $p\text{CO}_2$  over the coastal–open–ocean continuum (Figs. 2–4). However, the higher spatial resolution outperforms its lower resolution counterpart in reproducing fine-scale features of  $p\text{CO}_2$  in the transition from nearshore regions to the adjacent open ocean. The increase in model spatial  
 390 resolution translates into a greater spatial coverage of the continental shelves such as Labrador Sea, Northern Europe, and Sea of Japan (Fig. 3), and thus an increase in the number of data over the coastal domain. The increase in spatial resolution allows a gain in prediction probability of  $p\text{CO}_2$  variations on the order of roughly 2% over the Eastern Boundary Currents to 8% over the Western South Atlantic (Figs. 2–3b).



**Figure 5.** Time series of surface ocean  $p\text{CO}_2$  at coastal observing stations (Table A2 and Fig. A1b): model best estimate (curve),  $1\sigma$ -uncertainty (envelope), and monthly average of in situ observations (point). The reconstructed data at  $1^\circ$  (r100) and  $0.25^\circ$  (r025) resolutions are co-located to in situ observations provided by Sutton et al. (2019). Means of the best estimate and  $1\sigma$ -uncertainty ( $\mu \pm \sigma$ ) calculated over the observing time span are shown in brackets. Statistics include number of months with observations ( $N$ ), Bias, RMSD, and  $r^2$  computed for the two reconstructions.  $\sigma_{p\text{CO}_2}^t$  stands for temporal standard deviation from monthly averages of  $p\text{CO}_2$  observations.

Reconstruction skill of seasonal to inter-annual variability of  $p\text{CO}_2$  is further assessed at eight coastal monitoring sites (Sutton et al., 2019) and illustrated in Fig. 5 (see Sect. 2.2 and Table A2 for data description and Fig. A1b for station locations). The temporal variability of  $p\text{CO}_2$  reported for these time series sites reflects a combination of processes (Sutton et al., 2019), e.g., California Current System (CAPEARAGO and CCE2), western coastal upwelling (CAPEELIZABETH), eutrophica-



tion enhancing respiration of CO<sub>2</sub> (FIRSTLANDING), and multiple stressors on coral reef environments (CHEECAROCKS, GREYREEF). Results from reconstructions at two spatial resolutions are compared: 1° (100, black curve) and 0.25° (r025, color curve). As shown in Fig. 5 (scattered points for observations) time series of coastal pCO<sub>2</sub> are still short. The longest 400 time series covers 127 months of pCO<sub>2</sub> monitoring since 2010 (CCE2) while the shortest one contributes 17 months with observations (FIRSTLANDING).

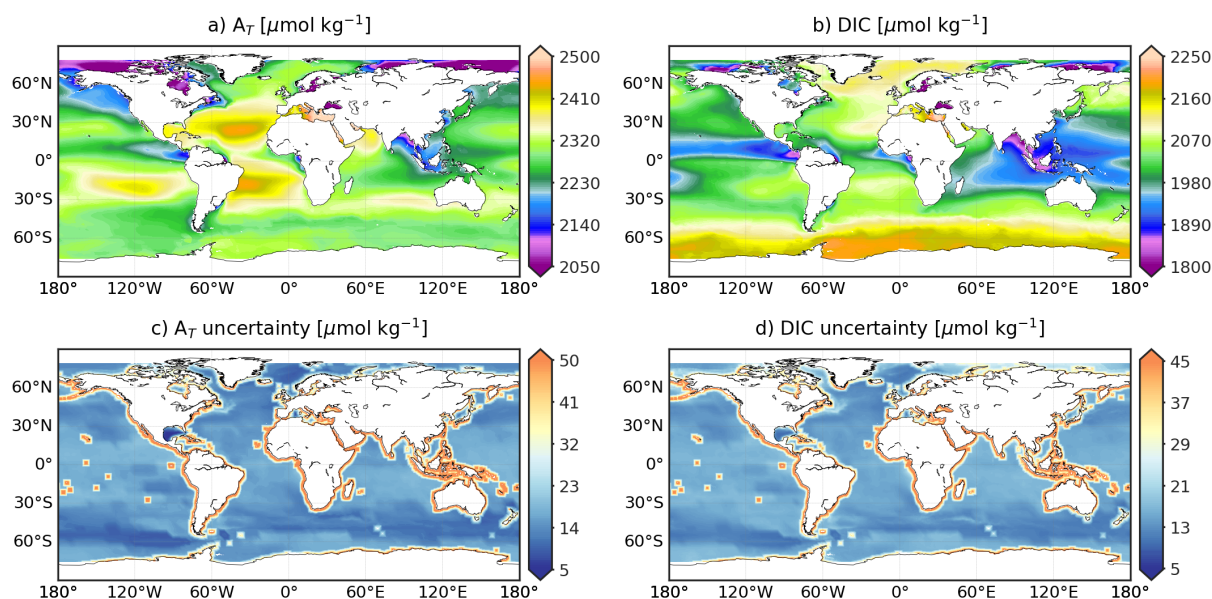
Analyzing the eight station time series, we have found that data have been sampled within a few days with an average offset of about a week from the month center. At these coastal sites, the temporal standard deviation from monthly averages of pCO<sub>2</sub> (σ<sup>t</sup><sub>pCO<sub>2</sub></sub>) exceeds measurement errors (2 μatm, Sutton et al., 2019). σ<sup>t</sup><sub>pCO<sub>2</sub></sub> ranges from 20.12 μatm at GREYREEFF to values 405 as large as 65.6 μatm at CAPEARAGO or 69.98 μatm at FIRSTLANDING. The monthly average of pCO<sub>2</sub> might not be adequately represented by discreet samples at sites with a large temporal standard deviation of pCO<sub>2</sub>. The misfit between the monthly reconstruction and discreet observations is exacerbated in dynamical coastal environments and might explain in part the large RMSD of reconstructions of monthly coastal pCO<sub>2</sub> (e.g., GREYREEFF: 38.34 μatm, CAPEARAGO: 79.86 μatm, 410 FIRSTLANDING: 77.32 μatm) for the r025 reconstruction. The RMSD is mostly lower for the FFNN reconstruction at 0.25° resolution compared to the FFNN at 1° resolution by 2.11 μatm (CCE2) to 23.32 μatm (COASTALMS). Similarly, r<sup>2</sup> increases between 7%-23% at higher resolution. Overall, seasonal to interannual variations of coastal-ocean pCO<sub>2</sub> are better reproduced in the reconstruction at 0.25° resolution (Fig. 5).

## 5.2 Total alkalinity and dissolved inorganic carbon

415 This section presents and analyzes global ocean surface reconstructions of total alkalinity (A<sub>T</sub>) and dissolved inorganic carbon (DIC) at monthly, 0.25° resolutions over 1985-2021. GLODAPv2.2022 bottle data (Sect. 2.2) serve as reference data for model evaluation. Model reconstruction skill is further assessed at the four Eulerian time series sites: BATS, DYFAMED, ESTOC, and HOTS (Table 3).

Figure 6 shows spatial distributions of the climatological mean and uncertainty (Eq. 5) for A<sub>T</sub> and DIC. Despite being in part 420 influenced by common biological and physical processes, both properties have contrasting distributions due to the strong correlation between surface ocean A<sub>T</sub> and salinity (Lee et al., 2006; Broullón et al., 2019), as well as the contribution of air-sea gas exchange and biological productivity on surface ocean DIC levels (Feely et al., 2001; Takahashi et al., 2014). Over subtropical Atlantic gyres and the Mediterranean Sea, oceanic areas with net evaporation, A<sub>T</sub> exceeds 2400 μmol kg<sup>-1</sup>. Total alkalinity falls below 2150 μmol kg<sup>-1</sup> in regions where precipitation, river freshwater runoff, or seasonal sea-ice melting dilute surface 425 water salinity (e.g., subpolar North Pacific, Arctic, and equatorial river outflows). The distribution of DIC is relatively uniform between the Atlantic, Pacific, and Indian Ocean basins, but shows pronounced latitudinal gradients. High concentrations of DIC are found throughout the Southern Ocean (DIC > 2100 μmol kg<sup>-1</sup>) where strong upwelling brings up subsurface water enriched in CO<sub>2</sub> and nutrients. The inefficient utilization of nutrients in this high nutrient low chlorophyll region limits the biological drawdown of DIC allowing the massive DIC input to be spread horizontally by westerlies (Key et al., 2004; Men- 430 viel et al., 2018). Levels of DIC below 1900 μmol kg<sup>-1</sup> are reconstructed over the Equatorial Pacific, the Equatorial Eastern Atlantic, the Eastern Indian Ocean, and coastal areas on the Arctic Ocean. While low DIC levels associated with Equatorial up-



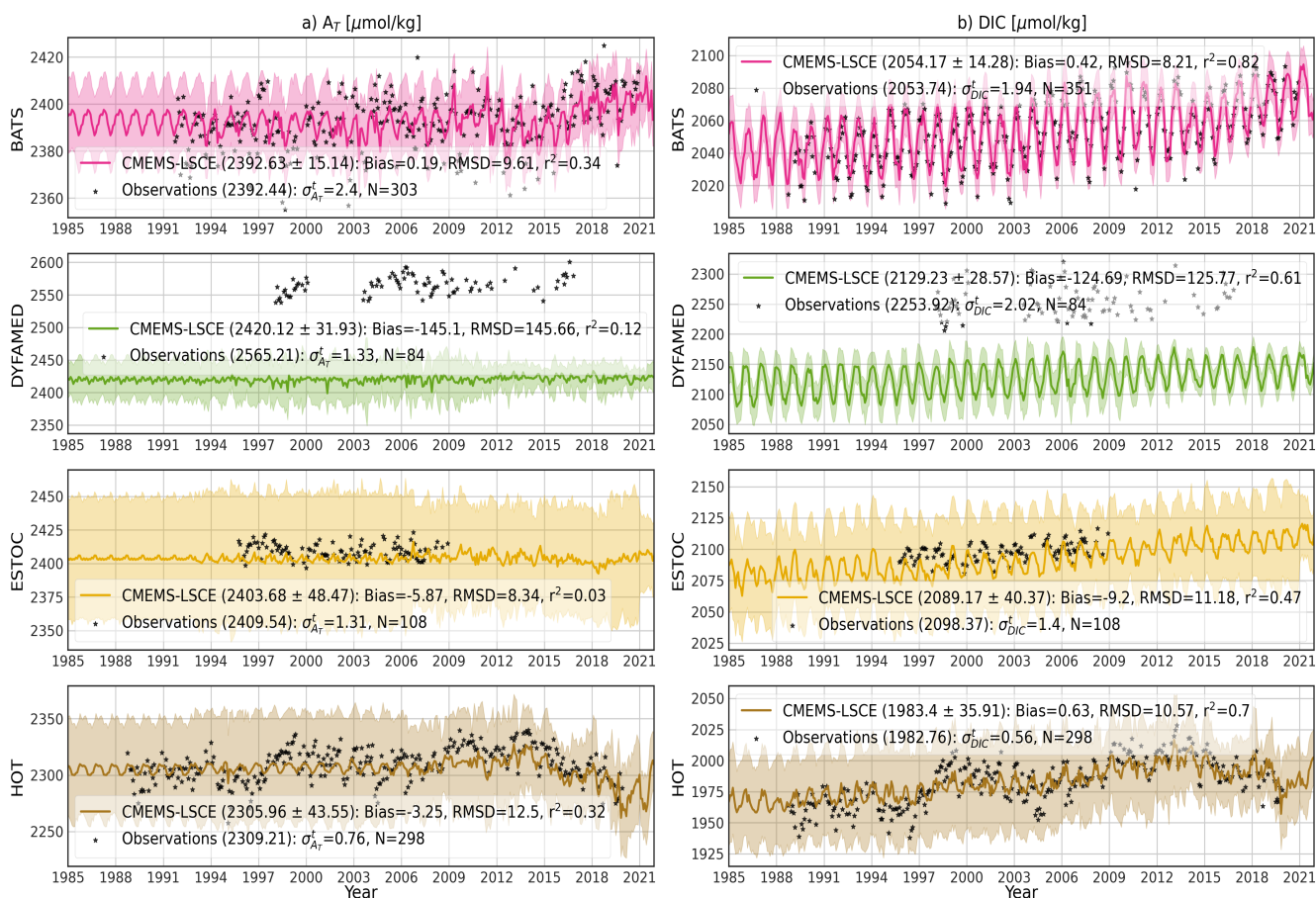


**Figure 6.** CMEMS-LSCE  $A_T$  and DIC over the global ocean at a spatial resolution of  $0.25^\circ$ . Temporal means of the model best estimate and  $1\sigma$ -uncertainty per grid cell over 1985–2021 are calculated by using Eq. (5).

welling reflect gas exchanges across the air-sea interface and enhanced biological production, the interaction between physical and biogeochemical processes at work in the Indian Ocean are less well understood (Takahashi et al., 2014). Low DIC levels found close to river mouths reflect outgassing of  $\text{CO}_2$  across the salinity gradient, as well as enhanced biological uptake fueled by river nutrient inputs. Representation uncertainty (Fig. 6-cd) associated with monthly alkalinity and DIC reconstructions is lower than  $20 \mu\text{mol kg}^{-1}$  throughout the open ocean. The open-ocean  $\sigma$ -to- $\mu$  ratio (Eq. 8) ranges between 0.5 – 1.5% which is relatively small (Fig. A9-cd). The largest model uncertainty ( $\sigma > 30 \mu\text{mol kg}^{-1}$ ) is computed nearshore and surrounding oceanic islands, a feature inherited from input uncertainty associated with the CMEMS salinity product (Fig. A8a).

We qualify monthly,  $0.25^\circ$  reconstructions of  $A_T$  and DIC with measurements from GLODAPv2.2022 (Lauvset et al., 2022) for the 37-year period (Table 5 and Fig. 8). The global open-ocean reconstruction scores a RMSD of  $22.09 \mu\text{mol kg}^{-1}$  and a  $r^2$  of 0.9 in  $A_T$ . Similar numbers are found for DIC (RMSD =  $22.67 \mu\text{mol kg}^{-1}$  and  $r^2 = 0.9$ ). The model scores the good fit in the open Indian Ocean with RMSD smaller than  $15.5 \mu\text{mol kg}^{-1}$  and  $r^2$  above 0.92 for both variables. The reconstruction deviates from GLODAP data in the western North Atlantic, subpolar North Pacific, tropics, and nearby major rivers (Fig. 8-abcd).

$A_T$  and DIC are underestimated in the continental shelves of north Alaska and the northeastern Atlantic, the Mediterranean Sea, South China Sea, and nearby river plumes (Fig. 8-ac). The Arctic yields the poorest estimations among all the ocean basins with a global RMSD over  $100 \mu\text{mol kg}^{-1}$  (Table 5). The prediction probability of variability in  $A_T$  [DIC] is relatively large for the open ocean 79% [71%], but rather unsatisfying over the coastal ocean (46% [40%]). Extrapolating these carbonate variables towards the shore remains challenging with much higher errors and uncertainty estimates obtained over the continental shelf

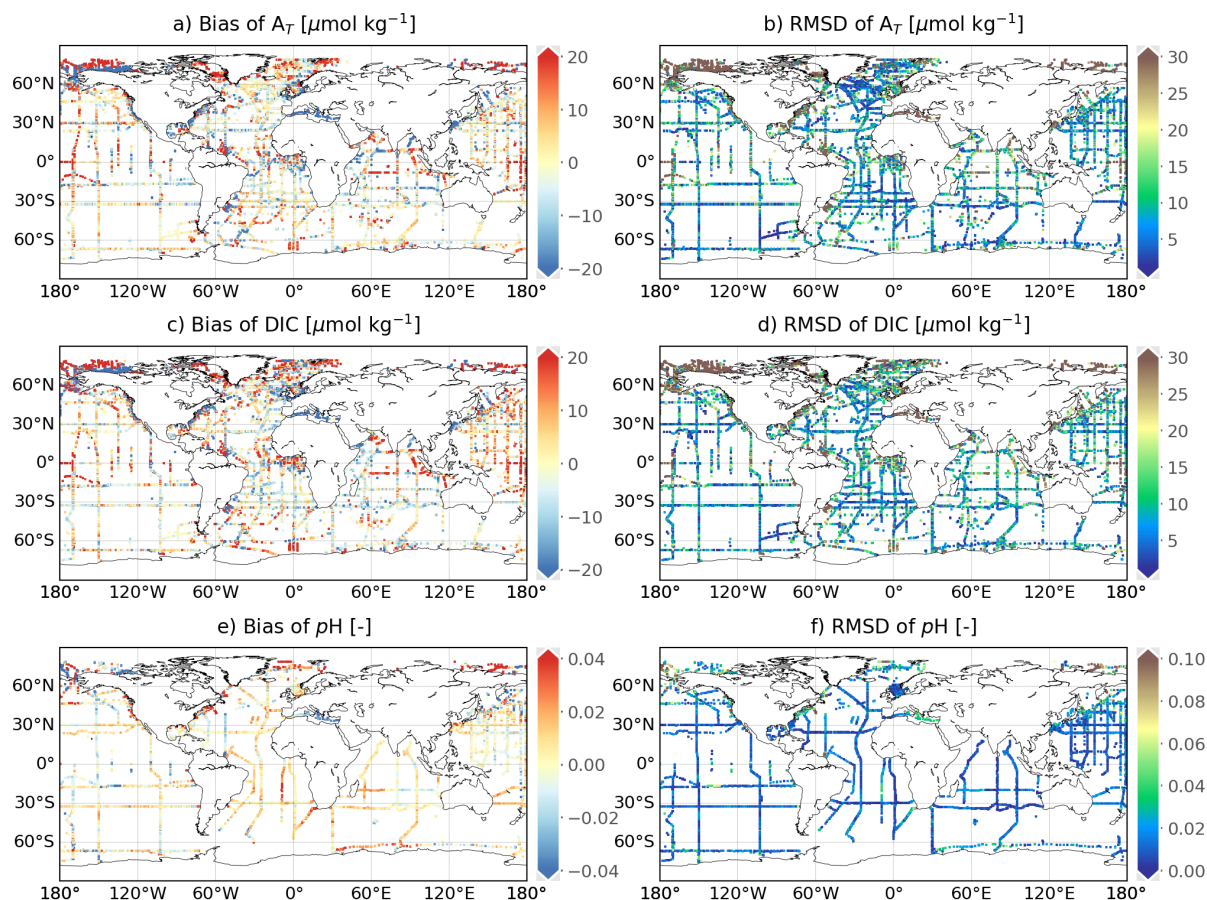


**Figure 7.** Monthly time series of  $A_T$  and DIC at BATS, DYFAMED, ESTOC and HOT stations (Table 3 and Fig. A1b): model best estimate (curve),  $1\sigma$ -uncertainty (envelope), and monthly average of surface (0-10 m) observations (point). Means of the best estimate and  $1\sigma$ -uncertainty ( $\mu \pm \sigma$ ) calculated over the observing time span are shown in brackets if accessible. Statistics include number of months with observations ( $N$ ), Bias, RMSD, and  $r^2$ .  $\sigma_{A_T}^t$  [ $\sigma_{DIC}^t$ ] stands for temporal standard deviation from monthly averages of  $A_T$  [DIC] observations.

450 compared to the open-ocean reconstruction (Table 5, Figs. 6-cd and . 8-abcd). The coastal-ocean errors are on the order of 10% of the global mean values of  $A_T$  and DIC (Table 7).

The reconstruction of  $A_T$  distributions relies on LIAR coefficients fit with GLODAPv2 data (Olsen et al., 2016) covering the years before 2015. These data are also part of the latest version GLODAPv2.2022 (Lauvset et al., 2022). They do therefore not correspond to an independent dataset for the evaluation data of the CMEMS-LSCE reconstruction. To overcome this limitation, reconstructions of  $A_T$  and DIC are compared to observations for Eulerian time series stations: BATS, DYFAMED, ESTOC, and HOT (see Table 3 and Fig. A1b for data sources and station locations). Figure 7 illustrates the comparison between monthly time series of  $A_T$  and DIC extracted from the CMEMS-LSCE datasets and measurements at these long-term monitoring sites. The four stations stand out as sustained long-term observation time series for carbonate system variables. More than 270 [80]





**Figure 8.** Spatial distribution of reconstruction skills for  $A_T$ , DIC, and  $pH$  over 1985-2021. Mean model-data difference (Bias) and root-of-mean square-deviation (RMSD) between the reconstruction and GLODAPv2.2022 surface data (0-10 m) at a spatial resolution of  $0.25^\circ$ . The size of grid cells is scaled upon a better visualization.

months in the years 1988-2021 [1995-2009 and 1998-2017] include measurements of  $A_T$  and DIC at BATS and HOT [ESTOC and DYFAMED]. The reconstructed time series fit monthly averages of in situ measurements well. Mean estimates of  $A_T$  [DIC] over the observing period are about 2305 [1983]  $\mu\text{mol kg}^{-1}$  at HOT to 2420 [2129]  $\mu\text{mol kg}^{-1}$  at DYFAMED. At all the stations (DYFAMED excepted) and for the two variables, model-observation misfit is small (Bias  $< 10 \mu\text{mol kg}^{-1}$ , RMSD  $< 13 \mu\text{mol kg}^{-1}$ ) relative to the aforementioned mean estimates. The highest offset between the CMEMS-LSCE estimation and observations for all the stations is found at DYFAMED ( $A_T$ :  $-145.1 \mu\text{mol kg}^{-1}$ , DIC:  $-124.69 \mu\text{mol kg}^{-1}$ ). DYFAMED provides long-term time series of  $A_T$  and DIC measurements in the Northwestern Mediterranean Sea (Fig. A1b). Salinity and alkalinity have substantial values due to the net evaporation (Coppola et al., 2020). The average of  $A_T$  in the Mediterranean Sea exceeds that for the global ocean by 10% (Palmiéri et al., 2015). These characteristics set the Mediterranean Sea aside from the ocean basins. LIAR (Carter et al., 2018) was trained on GLODAPv2 (Olsen et al., 2016) including a few observations in



**Table 5.** Skill scores computed between CMEMS-LSCE and GLODAPv2.v2022 in  $A_T$ , DIC, and  $pH$  over the period 1985-2021. Total numbers of data, RMSD (Eq. 10), and  $r^2$  (Eq. 11) are reported for both the open ocean (O) and coastal region (C).

Basins			Number of data		$A_T$ [ $\mu\text{mol kg}^{-1}$ ]		DIC [ $\mu\text{mol kg}^{-1}$ ]		pH [-]	
			$A_T$ -DIC	(pH)	RMSD	$r^2$	RMSD	$r^2$	RMSD	$r^2$
0.	Globe	(O)	10269	(5411)	22.09	0.90	22.67	0.90	0.022	0.70
		(C)	6309	(2080)	82.01	0.72	72.39	0.62	0.060	0.45
1.	Arctic	(O)	103	(26)	107.09	0.79	113.28	0.71	0.106	0.32
		(C)	1635	(300)	148.71	0.46	126.77	0.4	0.107	0.48
2.	Atlantic	(O)	2785	(932)	30.10	0.74	28.66	0.72	0.028	0.58
		(C)	2422	(941)	44.50	0.71	39.09	0.69	0.046	0.45
3.	Pacific	(O)	4539	(3222)	13.61	0.92	15.95	0.92	0.019	0.74
		(C)	1380	(639)	28.43	0.76	44.36	0.45	0.057	0.34
4.	Indian Ocean	(O)	1177	(551)	15.05	0.92	13.79	0.96	0.012	0.90
		(C)	328	(62)	16.56	0.92	21.97	0.90	0.013	0.82
5.	Southern Ocean	(O)	1665	(680)	10.96	0.64	13.21	0.92	0.019	0.68
		(C)	544	(138)	22.53	0.50	24.48	0.77	0.023	0.65

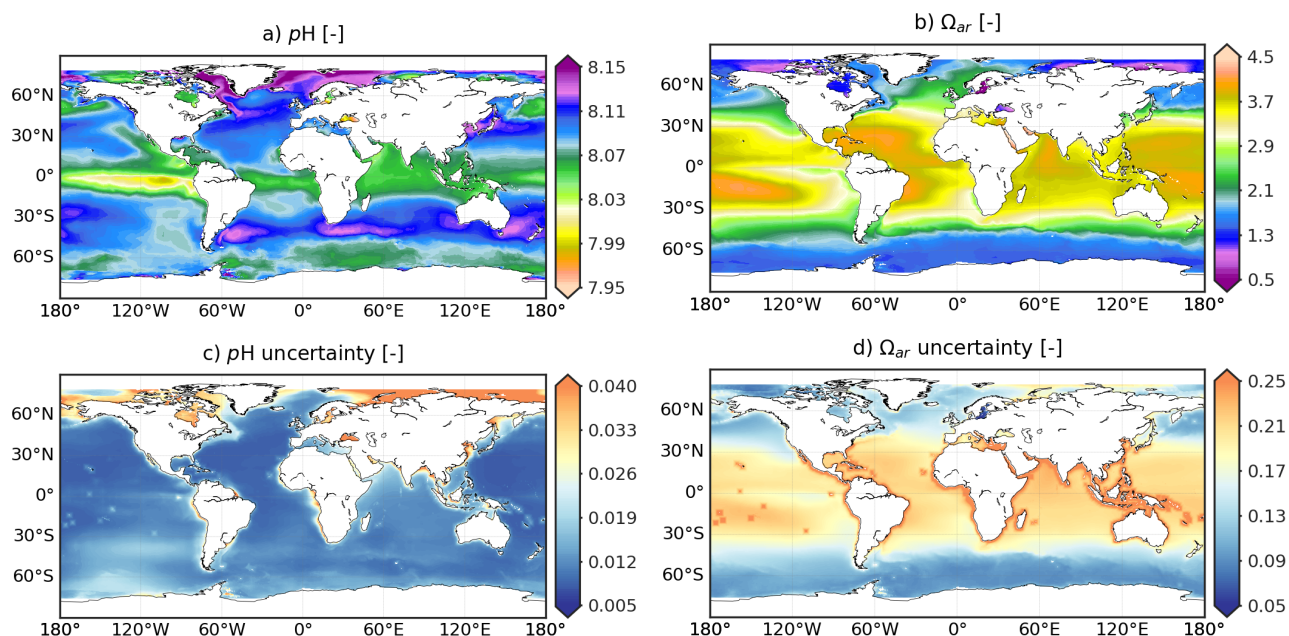
this area. The distinct relationship between alkalinity and salinity prevailing in the Mediterranean Sea is likely not reproduced by LIAR leading to an underestimation of  $A_T$  and a systematic bias to DIC at DYFAMED (Fig. 7). ESTOC is located close to the North Atlantic east coast and under the influence of the Canary Current System (CCS, Fig. A1). Spatial gradients and temporal variability are higher in the CCS (Fig. 2c) compared to BATS and HOT which are both located in the center of subtropical gyres. The lowest prediction skill of temporal variability is obtained for ESTOC. Particularly, seasonality to multi-year variations in DIC are predicted at  $r^2 = 0.47$  for ESTOC compared to  $r^2 > 0.7$  for BATS and HOT. Over all the stations, the model underestimates temporal changes of  $A_T$  (Fig. 7a; BATS:  $r^2 = 0.33$ , DYFAMED:  $r^2 = 0.12$ , ESTOC:  $r^2 = 0.03$ , HOT:  $r^2 = 0.32$ ) which can be attributed to the large discrepancy in variability between in situ measurements and the CMEMS time series of salinity (Fig. A10a; BATS:  $r^2 = 0.33$ , DYFAMED:  $r^2 = 0.19$ , ESTOC:  $r^2 = 0.03$ , HOT:  $r^2 = 0.35$ ). Model uncertainty ( $1\sigma$ -envelop) of monthly  $A_T$  and DIC estimates (Fig. 7a) is also inflated somewhat proportional to the CMEMS salinity product uncertainty (Fig. A10a).

### 5.3 Surface ocean $pH$ and saturation state with respect to carbonate minerals

Surface ocean  $pH$  and saturation states with respect to aragonite ( $\Omega_{ar}$ ) and calcite ( $\Omega_{ca}$ ) are critical indicators used to measure ocean acidification. This section first presents an overall evaluation of these variables. We then introduce estimates involved in the monitoring of ocean acidification in 1985-2021 as an essential application of the CMEMS-LSCE surface ocean carbon product.



### 5.3.1 General analysis and evaluation



**Figure 9.** CMEMS-LSCE  $pH$  and  $\Omega_{ar}$  over the global ocean at a spatial resolution of  $0.25^\circ$ . Temporal means of the model best estimate and  $1\sigma$ -uncertainty per grid cell over 1985–2021 are calculated by using Eq. (5).

485 The spatial distribution of surface ocean  $pH$  reported on total hydrogen ion ( $H^+$ ) scale is shown in Fig. 9 (the corresponding figure for  $H^+$ , Fig. A4, is included in the supplementary). Both temporal means of the best model estimate and  $1\sigma$ -uncertainty of  $pH$  share spatial patterns with  $pCO_2$  (Fig. 1). Variables  $pH$  and  $pCO_2$  correlate closely through equilibrium relationships of dissolved  $CO_2$  in seawater: an increase in  $pCO_2$  generally corresponds to a decrease in  $pH$ . The distribution of the climatological mean of  $pH$  displays a gradient with latitude between 8.03 and 8.11  $pH$  units across most of the basins (Fig. 9a). Values of  $pH$  below 8 are associated with the upwelling of  $CO_2$ -rich waters (e.g., Eastern Equatorial Pacific, Western Arabian Sea).  $pH$  exceeds 8.15 in sub- and polar cold surface water and in the regions with high biological productivity (e.g. Labrador Sea, Nordic Seas, Southern Ocean between  $35^\circ S$ – $50^\circ S$ ).

495 The saturation state of surface ocean waters with respect to calcium carbonate minerals aragonite and calcite is defined as the ratio of the product of the concentrations of calcium ions ( $Ca^{2+}$ ) and carbonate ions ( $CO_3^{2-}$ ) to the solubility of the respective calcium carbonate mineral ( $CaCO_3$ ) in surface seawater (Eq. A5). Aragonite being the more soluble polymorph, its degree of saturation ( $\Omega_{ar}$ ) is smaller than that of calcite ( $\Omega_{ca}$ ) (Mucci, 1983). With the exception of this offset, the spatial distributions of their climatological means share common spatial patterns over the global ocean (Figs. 9b and A5a). Surface seawater is generally supersaturated, i.e.,  $\Omega_{ar}$  and  $\Omega_{ca}$  greater than 1. The magnitude of surface ocean calcium carbonate saturation state varies with latitude. Values as large as 3.7–4.5 [5–7] for aragonite [calcite] are reconstructed in subtropical and tropical regions.



500  $\Omega_{ar}$  and  $\Omega_{ca}$  decrease toward the poles. In the Southern Ocean, surface seawater enriched in  $\text{CO}_2$  from vertical mixing has  $\Omega_{ar}$  [ $\Omega_{ca}$ ] values in the range of 1.5-2.1 [2-3.4]. Low saturation states are also computed in the Arctic and for waters of upwelling regimes (Fig. 9b). Locally  $\Omega_{ar}$  drops below 1.3, and even fall under the  $\text{CaCO}_3$  dissolution threshold of 1 (Gattuso and Hansson, 2011) in the Arctic water runoff and Baltic sea.

The uncertainty ( $1\sigma$ ) of  $p\text{H}$ ,  $\Omega_{ar}$ , and  $\Omega_{ca}$  propagated the speciation of the  $\text{CO}_2$  system takes into account the ensemble spread of  $p\text{CO}_2$  estimates and analysis errors of other variables (Sect. 3.3). Monthly  $p\text{H}$  uncertainty estimates fall in the 95% confidence interval of [0.008, 0.036] with a global mean value of 0.011. These estimates are in close agreement with the global uncertainty between 0.01-0.022  $p\text{H}$  units calculated by Jiang et al. (2019), Iida et al. (2021), and Gregor and Gruber (2021).  $p\text{H}$  uncertainty is typically larger than 0.03 in the Arctic and in coastal regions (Figs. 9c). In contrast, the reconstructions of  $\Omega_{ar}$  and  $\Omega_{ca}$  are subject to high uncertainty ( $\sigma > 0.175$ ) between  $30^\circ\text{S}$ - $30^\circ\text{N}$  (Fig. 9d and A5b). Regarding the  $\sigma$ -to- $\mu$  ratio, mean uncertainty estimates per cell for the saturation states in the (sub-) tropical band are relatively small compared to the mean of the best monthly estimates (Figs. A9-ef). The Arctic and the coastal oceans remain the regions with largest reconstruction uncertainties for  $\Omega_{ar}$  and  $\Omega_{ca}$ , as well as for  $p\text{CO}_2$  and  $p\text{H}$  (Figs. A9-ab). Excluding these regions,  $R(\sigma, \mu)$  (Eq. 8) is less than 0.3% for  $p\text{H}$  and 8% for  $\Omega_{ar}$  and  $\Omega_{ca}$ .

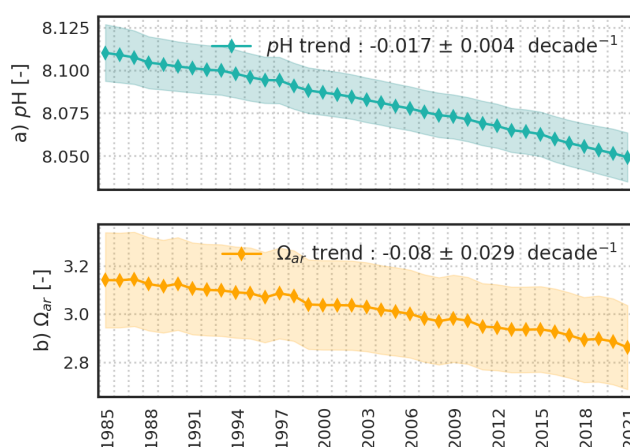
The monthly CMEMS-LSCE reconstruction at  $0.25^\circ$  resolution is assessed against  $p\text{H}$  measurements from GLODAPv2.2022 bottle data (Table 3). For the period 1985-2021, the global RMSD amounts to 0.022 [0.060]  $p\text{H}$  units and  $r^2$  scores at 0.70 [0.45] over the open [coastal] ocean (Table 5). Model bias lies within  $[-0.01, 0.01]$   $p\text{H}$  units and RMSD is below 0.02  $p\text{H}$  units over the open ocean, except for high latitudes over  $60^\circ$  (Figs. 8-ef). At local scale, the eight coastal time series from Sutton et al. (2019) are used for further evaluation. There exists much less evaluation data for  $p\text{H}$  than for  $p\text{CO}_2$ , e.g., only 2 months of monitoring  $p\text{H}$  at COASTALLA and FIRSTLANDING (Table A3). Monthly time series of CMEMS-LSCE  $p\text{H}$  are coherent with these  $p\text{H}$  measurements (Tables A2 and A3). Measurement uncertainty of  $p\text{H}$  at these coastal sites is reported to be around 0.02  $p\text{H}$  units. RMSD can be as small as 0.035 and 0.04  $p\text{H}$  units at CCE2 and GRAYSREEF while it is over 0.05  $p\text{H}$  units at the other stations (e.g., COASTALLA: 0.068, CAPEARAGO: 0.069). Similar to the  $p\text{CO}_2$  time series (Fig. 5, Sect. 5.1),  $p\text{H}$  has been monitored with low sampling frequency (roughly a few days in the tracking month) and the temporal sampling deviation of instantaneous observations from monthly averages ( $\sigma_{p\text{H}}^t$ ) is significant. This temporal sampling uncertainty of  $p\text{H}$  contributes to the mismatch between model estimates and observations. For example,  $\sigma_{p\text{H}}^t$  amounts to 0.048  $p\text{H}$  units at CCE2 and 0.020  $p\text{H}$  units at GRAYSREEF, and reaches highest values of 0.078  $p\text{H}$  units at COASTALLA and 0.086  $p\text{H}$  units at CAPEARAGO. Although model-observation misfit and model uncertainty remain high over the coastal sector (see also Figs. 8-ef and 9c), their estimates do not surpass 1% of the global mean  $p\text{H}$  (8.082). The reconstructed  $p\text{H}$  time series reproduce measurement variability with relatively high correlation,  $r^2$  in [0.21, 0.69], that reinforces the reliability of CMEMS-LSCE  $p\text{H}$  data.

### 5.3.2 Ocean acidification: key features from global to local scales

The monthly,  $0.25^\circ$  CMEMS-LSCE datasets of  $p\text{H}$ ,  $\Omega_{ar}$ , and  $\Omega_{ca}$  are at the basis of of two CMEMS ocean indicators monitoring surface ocean acidification from 1985 to 2021: (1) annual global means and (2) global trend maps.



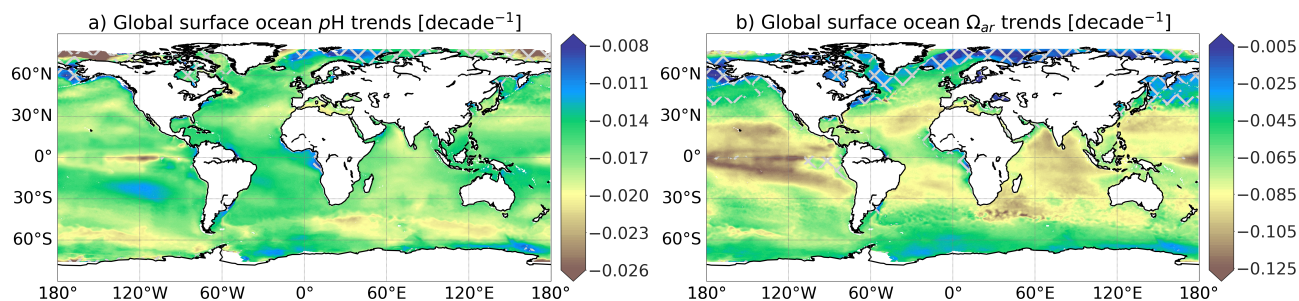
In Fig. 10, we present annual global means of surface ocean  $pH$  and saturation states for aragonite ( $\Omega_{ar}$ ). An illustration of calcite ( $\Omega_{ca}$ ) is provided in the Appendix (Fig. A13a). For each variable, the calculation of annual global area-weighted means of best estimates (line) and  $1\sigma$ -uncertainties (envelope) follows Eq. (7). The trends reported in the legend result from linear least-squares regression on annual global means of 100-ensembles of the carbonate system variables. These ensembles are generated with Gaussian distribution having the mean and variance as best model estimate  $\mu$  and squared uncertainty ( $\sigma^2$ ) at monthly time steps and  $0.25^\circ$ -grid cells, respectively.  $pH$  decreases from  $8.110 \pm 0.017$  in 1985 to  $8.049 \pm 0.014$  in 2021 with a descend rate of  $-0.017 \pm 0.004 \text{ decade}^{-1}$ . Similar trends are found for the surface ocean saturation states with respect to calcium carbonate minerals. The global mean estimates of  $\Omega_{ar}$  [ $\Omega_{ca}$ ] amount to  $3.141 \pm 0.198$  [ $4.807 \pm 0.302$ ] and  $2.862 \pm 0.174$  [ $4.372 \pm 0.266$ ] for the open and coastal oceans. The saturation state declines at a rate of  $-0.080 \pm 0.029 \text{ decade}^{-1}$  with respect to aragonite while the reduction is steeper for calcite ( $-0.114 \pm 0.045 \text{ decade}^{-1}$ ).



**Figure 10.** Yearly global area-weighted mean of surface seawater  $pH$  reported on total scale (a) and surface ocean saturation states with respect to aragonite (b). Global means of the best estimate ( $\mu$ , plain line) and of uncertainty ( $\sigma$ , envelop) are computed with Eq. (7a). Trend and uncertainty in the legend are computed with linear regressions on the 100-member ensemble of yearly global means for each variable.

Global trend maps of surface ocean  $pH$ ,  $\Omega_{ar}$ , and  $\Omega_{ca}$  over the entire period are illustrated in Figs. 11 and A13b. Linear least-squares regression is used to estimate secular trends at every  $0.25^\circ$ -grid cell. The linear fits of each variable against time rely on the 100-member ensemble generated with the best estimates and propagated uncertainties of  $pH$ ,  $\Omega_{ar}$ , and  $\Omega_{ca}$  (see Figs. A14 for examples). Regression slope and residual standard deviation estimates are defined as linear trend and uncertainty of  $pH$ ,  $\Omega_{ar}$ , and  $\Omega_{ca}$ . Hatched area represents  $pH$  [ $\Omega_{ar}$  and  $\Omega_{ca}$ ] trend estimates ( $\mu$ ) with highest uncertainties ( $\sigma$ ), i.e.,  $\sigma$ -to- $\mu$  ratio (Eq. 8) above 10% [20%]. These regions include a portion of the Arctic, Antarctic, equatorial Pacific, and coastal ocean (Figs. 11, A11, and A12). 95% of  $pH$  trend estimates over the global ocean is in the range of  $[-0.022, -0.012] \text{ decade}^{-1}$  (Fig. 11a). In the broad open ocean of the tropics and subtropics,  $pH$  has been declining around  $-0.018 \text{ decade}^{-1}$  to  $-0.012 \text{ decade}^{-1}$ . Faster decrease rates are found in the Indian Ocean and Southern Ocean with values between  $-0.022$  and  $-0.018 \text{ decade}^{-1}$ . Fastest reductions are computed for the eastern equatorial Pacific and the Arctic with rates exceeding  $-0.025 \text{ decade}^{-1}$ . A





**Figure 11.** Global trend maps of surface seawater  $pH$  reported on total scale (a) and surface ocean saturation states with respect to aragonite (b). Linear trend of CMEMS-LSCE  $pH$  and  $\Omega_{ar}$  is estimated per  $0.25^\circ$ -grid cell over 1985–2021. Cross-hatching covers the regions where uncertainty over 10% [20%] of  $pH$  [ $\Omega_{ar}$ ] trend estimates.

similar magnitude of  $pH$  trends over these regions is also found in (Lauvset et al., 2015; Leseurre et al., 2022; Ma et al., 555 2023). The spatial distribution of saturation states with respect to calcium carbonate minerals generally shows the opposite latitudinal pattern (Figs. 11b and A13b). The magnitude of  $\Omega_{ar}$  [ $\Omega_{ca}$ ] trends over the  $30^\circ S$ – $30^\circ N$  band can be as large as  $-0.086 \text{ decade}^{-1}$  [ $-0.134 \text{ decade}^{-1}$ ] to the greatest extent of  $-0.186 \text{ decade}^{-1}$  [ $-0.275 \text{ decade}^{-1}$ ] (e.g., eastern equatorial Pacific). Trends of  $\Omega_{ar}$  and  $\Omega_{ca}$  computed in polar and subpolar northern hemisphere regions are not significant.

Trend estimates derived from reconstructions of  $pH$  and  $\Omega_{ar}$  are evaluated at seven time series stations (Bates et al., 2014) 560 in Table 6. Time series locations are shown in Fig. A1b. With the exception of CARIACO and HOT excepted for which  $pH$  measurements are available, long-term trend estimates by Bates et al. (2014) rely on time series of  $pH$  and  $\Omega_{ar}$  calculated via speciation from measurements of  $A_T$  and DIC. A 100-member ensemble of monthly time series of  $pH$  and  $\Omega_{ar}$  are extracted from the  $0.25^\circ$ -grid box nearest to each monitoring station. Linear least-squares regression is then used to infer estimates of their secular trends and associated uncertainties (see Fig. A14 for illustration). Trend estimates derived from CMEMS-LSCE 565 reconstructions at HOT, BATS, ESTOC, and Munida are in line with previous studies for both  $pH$  and  $\Omega_{ar}$  (Dore et al., 2009; González-Dávila and Santana-Casiano, 2009; Bates et al., 2014). The magnitude of the trend estimate at Irminger Sea for 1985–2012 ( $pH$ :  $-0.014 \pm 0.001 \text{ decade}^{-1}$ ,  $\Omega_{ar}$ :  $-0.006 \pm 0.011 \text{ decade}^{-1}$ ) is smaller than that determined by Bates et al. (2014). However, the CMEMS-LSCE  $pH$  trend is consistent with the estimate by Pérez et al. (2021) ( $-0.017 \pm 0.002 \text{ decade}^{-1}$ ). Moreover,  $1\sigma$ -uncertainty reported for both  $pH$  and  $\Omega_{ar}$  trend estimates by Bates et al. (2014) is large at this station ( $pH$ : 570  $-0.025 \pm 0.006 \text{ decade}^{-1}$ ,  $\Omega_{ar}$ :  $-0.080 \pm 0.040 \text{ decade}^{-1}$ ) highlighting the associated uncertainty. Long-term trends of  $pH$  and  $\Omega_{ar}$  are also under-estimated at the Iceland Sea monitoring site, but the bias is not as large as at Irminger Sea (Table 6). Low data sampling frequency at these two stations (Table 1, Bates et al., 2014) could be on account of trend estimate deviation. At CARIACO, the CMEMS-LSCE time series yields a decrease in  $\Omega_{ar}$  of  $-0.059 \pm 0.053 \text{ decade}^{-1}$ , relatively close to Bates et al. (2014) ( $-0.066 \pm 0.028 \text{ decade}^{-1}$ ). The decrease in  $pH$  derived from CMEMS-LSCE is, however, larger than in Bates 575 et al. (2014).





**Table 6.** Secular trend estimates of  $pH$  and  $\Omega_{ar}$  at seven time-series stations (Bates et al., 2014). Trend and uncertainty estimates are reported as  $\mu \pm \sigma$ . Monthly time series in the CMEMS-LSCE datasets are extracted at the grid box nearest to each station location (Fig. A1b). For the first three stations, this study calculates linear trends starting in the year 1985. Brackets show values computed over the full period 1985-2021.

Stations	Coordinates	Time span	$pH$ trend [decade <sup>-1</sup> ]		$\Omega_{ar}$ trend [decade <sup>-1</sup> ]	
			Bates et al. (2014)	This study	Bates et al. (2014)	This study
1. Iceland Sea	68.00°N 12.66°W	1983-2012	-0.014 ± 0.005	-0.010 ± 0.001 (-0.014 ± 0.001)	-0.018 ± 0.027	-0.013 ± 0.011 (-0.025 ± 0.008)
2. Irminger Sea	64.30°N 28.00°W	1983-2012	-0.026 ± 0.006	-0.014 ± 0.001 (-0.016 ± 0.001)	-0.080 ± 0.040	-0.006 ± 0.011 (-0.039 ± 0.009)
3. BATS	32.00°N 64.00°W	1983-2012	-0.017 ± 0.001	-0.014 ± 0.001 (-0.016 ± 0.001)	-0.095 ± 0.007	-0.079 ± 0.016 (-0.074 ± 0.010)
4. ESTOC	29.04°N 15.50°W	1995-2012	-0.018 ± 0.002	-0.018 ± 0.002 (-0.019 ± 0.001)	-0.115 ± 0.023	-0.103 ± 0.031 (-0.089 ± 0.011)
5. HOT*	22.75°N 158.00°W	1988-2012	-0.016 ± 0.001	-0.016 ± 0.001 (-0.019 ± 0.001)	-0.084 ± 0.011	-0.100 ± 0.020 (-0.102 ± 0.011)
6. CARIACO*	10.50°N 64.66°W	1995-2012	-0.025 ± 0.004	-0.017 ± 0.003 (-0.018 ± 0.001)	-0.066 ± 0.028	-0.059 ± 0.053 (-0.099 ± 0.018)
7. Munida	45.70°S 171.50°E	1998-2012	-0.013 ± 0.003	-0.017 ± 0.002 (-0.017 ± 0.001)	-0.085 ± 0.026	-0.088 ± 0.032 (-0.070 ± 0.009)

\*Stations with direct observations of  $pH$ .

## 6 Conclusions and Discussion

This study presents the CMEMS-LSCE product, a dataset of six carbonate system variables (Table 1) covering the global surface ocean at a spatial resolution of  $0.25^\circ$  for every month in the period 1985-2021 (444 months). Datasets of individual carbonate system variables are built on the combination of the three methods. First, we adapt an ensemble of 100 feed-forward neural network models (CMEMS-LSCE-FFNN, Chau et al., 2022b) to estimate surface ocean partial pressure of  $CO_2$  ( $pCO_2$ ) at the pre-defined data resolution. Second, the high-resolution total alkalinity ( $A_T$ ) reconstruction is obtained by using locally interpolated alkalinity regression (LIAR, Carter et al., 2016, 2018). Finally, surface ocean  $pH$ , total dissolved inorganic carbon (DIC), and saturation states with respect to aragonite ( $\Omega_{ar}$ ) and calcite ( $\Omega_{ca}$ ) are calculated with the carbonate system speciation software (CO2SYS, Lewis and Wallace, 1998; Van Heuven et al., 2011; Orr et al., 2018), given the global monthly reconstructions of  $pCO_2$  and  $A_T$  and other environmental input data (Sect. 3). Results are 2D-fields of the best estimate and associated uncertainty ( $1\sigma$ ) of carbonate system variables available at each grid box of  $1\text{month} \times 0.25^\circ \times 0.25^\circ$ .  $1\sigma$ -uncertainty is referred to as the ensemble standard deviation of 100 FFNN outputs for  $pCO_2$  while it is propagated through the processing chain of LIAR and CO2SYS taking into account different uncertainty sources of input parameters for other variables.



Multiple observation-based datasets, which are not used for the CMEMS-LSCE reconstructions at monthly and 0.25° resolu-  
590 tions, serve as benchmarks in the assessments of product quality from global to local scales (e.g., Tables 4, 5, and A3; Figs. 2-5  
and 7-8). A summary of the primary statistics for all the six carbonate variables is presented Table 7. Over the full period  
1985-2021, CMEMS-LSCE yields global RMSDs of 14.29  $\mu\text{atm}$  and 27.55  $\mu\text{atm}$  in comparison with SOCATv2022  $p\text{CO}_2$  for  
the open and coastal oceans, respectively. Temporal variability of observation-based data is well reproduced with  $r^2$  of 0.83  
for the open ocean and 0.74 for the coastal domain. In comparison to CMEMS-LSCE at monthly and 1° resolutions (Chau  
595 et al., 2022b), the reconstructions over coastal areas are improved at higher resolution (Figs. 2-4). Furthermore, the monthly,  
0.25° reconstruction outperforms its 1° counterpart in reproducing horizontal and temporal gradients of  $p\text{CO}_2$  over a variety of  
oceanic regions as well as at nearshore time series stations (Figs. 2-5). Evaluations with GLODAPv2022 bottle data and time  
series stations results in good reconstruction skills for  $A_T$ , DIC, and  $p\text{H}$  at monthly and 0.25° resolutions (Tables 5 and A3,  
Figs. 7 and 8). At the global scale, the open-ocean reconstruction scores a RMSD smaller than 23  $\mu\text{mol kg}^{-1}$  and a  $r^2$  of 0.9 in  
600  $A_T$  and DIC. The model-observation deviation is higher in the coastal zone. However, it does not exceed 5% of the global mean  
values and  $r^2$  is above 0.6 for both coastal  $A_T$  and DIC. Regarding  $p\text{H}$ , the CMEMS-LSCE reconstruction provides estimates  
with RMSD= 0.022 [0.060] and  $r^2 = 0.7$  [0.45] over the open [coastal] ocean. From the statistics in Tables 4 and 5, the Indian  
Ocean and the Southern Ocean have poor data density (Fig. A2) but generally show the best global reconstruction among the  
ocean basins. Thus, model evaluation with different numbers of observation data might not reflect a fair comparison of skill  
605 scores (e.g., RMSD and  $r^2$ ) between regions. Data density is much higher in the Arctic, Atlantic, and Pacific than in the Indian  
and Southern Oceans. The increased data density reveals stronger spatio-temporal variability, for instance, related to coastal  
dynamics or upwelling than resolved in the two latter basins. RMSD and  $r^2$  computed on the lower data variability result in  
better model scores.

The spatial distribution of long-term mean  $1\sigma$ -uncertainty estimates (Figs. 1b, 6cd, and 9cd) indicates higher confidence  
610 levels for open-ocean estimates than over the coastal sector. The evaluation of temporal mean  $1\sigma$ -uncertainty estimates relative  
to climatological mean values  $\mu$  (Figs. 1a, 6ab, and 9ab) results in  $\sigma$ -to- $\mu$  ratio (Eq. 8) below 5% and 8% for  $p\text{CO}_2$  and  $\Omega_{ar}$ ,  
2% for  $A_T$  and DIC, and 0.4% for  $p\text{H}$  over the open ocean (Fig. A9). The  $\sigma$ -to- $\mu$  ratio reaches values as high as 10% to 20%  
for  $p\text{CO}_2$  and  $\Omega_{ar}$  in the coastal domain. The global mean of open-ocean  $1\sigma$ -uncertainty estimates (Eq. 7a) for CMEMS-LSCE  
 $p\text{CO}_2$  (8.48  $\mu\text{atm}$ ),  $A_T$  (16.66  $\mu\text{mol kg}^{-1}$ ), DIC (15.75  $\mu\text{mol kg}^{-1}$ ),  $p\text{H}$  (0.011), and  $\Omega_{ar}$  (0.180) are in line with those reported  
615 by previous studies despite being derived from different statistics. For instance, Iida et al. (2021) calculated  $1\sigma$ -uncertainty  
based on the median absolute deviation of regression model fits from open-ocean observations. Their approach yielded global  
 $\sigma$ -averages of 17.8  $\mu\text{atm}$ , 11.5  $\mu\text{mol kg}^{-1}$ , 0.018, and 0.110 for  $p\text{CO}_2$ , normalized DIC,  $p\text{H}$ , and  $\Omega_{ar}$ , respectively. In Gregor  
and Gruber (2021), the authors propagated the sum squared errors (global RMSD and measurement uncertainties) of  $p\text{CO}_2$   
(15  $\mu\text{atm}$ ) and  $A_T$  (22  $\mu\text{mol kg}^{-1}$ ) obtaining global uncertainty estimates of 19  $\mu\text{mol kg}^{-1}$  in DIC and 0.022 in  $p\text{H}$ . Mean  
620 uncertainty estimates over the coastal region are on the order of twofold that computed for the open ocean for these four  
variables (Table 7), corroborating results by Gregor and Gruber (2021) (Fig. 7).

Our high-resolution carbon data product opens the door to various analyses of the marine carbonate system from global to lo-  
cal scale. This study exemplifies an application of the data for monitoring ocean acidification over recent years. The monitoring



**Table 7.** Summary in global evaluation statistics for CMEMS-LSCE surface ocean carbonate system datasets at monthly,  $0.25^\circ$  resolutions over the period 1985–2021.  $\mu$  and  $\sigma$  stand for the global area-weighted means of monthly best estimates and  $1\sigma$ -uncertainties for each variable (Eq. 7). RMSD (Eq. 10) and  $r^2$  (Eq. 11) are computed with SOCATv2022 for  $p\text{CO}_2$  and GLODAPv2.2022 for  $p\text{H}$ ,  $A_T$ , and DIC. The division between the coastal (C) and open (O) oceans is at 400 km on a distance from the shore line (Fig. A1a).

Variables	Units	Sector	$\mu$	$\sigma$	RMSD	$r^2$
1. $p\text{CO}_2$	$\mu\text{atm}$	(O)	364.48	8.48	14.29	0.83
		(C)	359.35	17.10	27.55	0.74
2. $A_T$	$\mu\text{mol kg}^{-1}$	(O)	2305.78	16.66	22.09	0.90
		(C)	2263.02	38.36	82.01	0.72
3. DIC	$\mu\text{mol kg}^{-1}$	(O)	2031.12	15.75	22.67	0.90
		(C)	2008.65	33.40	72.39	0.62
4. $p\text{H}$	-	(O)	8.082	0.011	0.022	0.70
		(C)	8.082	0.021	0.060	0.45
5. $\Omega_{ar}$	-	(O)	3.059	0.180	-	-
		(C)	2.864	0.206	-	-
6. $\Omega_{ca}$	-	(O)	4.674	0.275	-	-
		(C)	4.384	0.314	-	-

indicators derived from the monthly,  $0.25^\circ$  surface ocean CMEMS-LSCE product consist of (1) yearly global means of surface  
 625 ocean  $p\text{H}$  and saturation states with respect aragonite  $\Omega_{ar}$  and calcite  $\Omega_{ca}$  and (2) global maps of multi-annual trends of surface  
 ocean  $p\text{H}$ ,  $\Omega_{ar}$ , and  $\Omega_{ca}$  (Figs. 10, 11, and A13). In 1985, the global mean surface ocean  $p\text{H}$  was  $8.110 \pm 0.017$ . It was  $8.049 \pm$   
 $0.014$  in 2021 (Fig. 10a). Over the same 37-year time period,  $\Omega_{ar}$  decreased from  $3.141 \pm 0.198$  to  $2.862 \pm 0.174$  (Fig. 10b).  
 The rate of decline of surface ocean  $p\text{H}$  and  $\Omega_{ar}$  was respectively  $-0.017 \pm 0.004 \text{ decade}^{-1}$  and  $-0.080 \pm 0.029 \text{ decade}^{-1}$   
 since 1985 (see also results for  $\Omega_{ca}$  in Sect. 5.3.2). Estimates of  $p\text{H}$  trend lie between  $[-0.022, -0.012] \text{ decade}^{-1}$  across most  
 630 of the open ocean (Fig. 11a). In general, surface ocean  $p\text{H}$  decreased more rapidly in the Indian Ocean and Southern Ocean  
 than the tropics and subtropics. These findings are in close agreement with the suggestions by Lauvset et al. (2015) and Ma  
 et al. (2023) but future studies would need to include analyses of underlying drivers to provide insight into regional differences  
 in  $p\text{H}$  changes. By contrast, the greatest reduction in surface ocean saturation states (Fig. 11b) was computed for the two latter  
 regions. The global trend maps of  $p\text{H}$  and  $\Omega_{ar}$  highlight the Eastern Equatorial Pacific as one of the vulnerable regions with  
 635 respect to ocean acidification. In this area, the decline rate of  $p\text{H}$  exceeds  $-0.025 \text{ decade}^{-1}$  and  $-0.186 \text{ decade}^{-1}$  for  $\Omega_{ar}$ .  
 The comparison of multi-annual trends of  $p\text{H}$  and  $\Omega_{ar}$  at time series stations (Table 6 and Fig. A14) highlighted the consistency  
 between CMEMS-LSCE estimates and previous studies (Dore et al., 2009; González-Dávila and Santana-Casiano, 2009; Bates  
 et al., 2014; Pérez et al., 2021). For most of these sites, the trends evaluated for 1985–2021 are greater than those relative to the  
 sub-period before the year 2012. The faster rate of ocean acidification over the full period compared to the pre-2012 probably  
 640 reflects a steeper acceleration in ocean uptake of anthropogenic  $\text{CO}_2$  in the last decade.

The production chain of CMEMS-LSCE carbonate system variables will be maintained and further improvements with  
 the aim to reduce model-observation misfit and improve the quantification of model uncertainty are on the way forward.  
 Being at the core of the chain, model upgrades of CMEMS-LSCE-FFNN will be tackled first. At the time, SOCAT does not  
 provide open-ocean data of  $\text{CO}_2$  fugacity gridded at monthly,  $0.25^\circ$  resolutions. Our ensemble-based approach draws thus



645 on two SOCATv2022 data sources: a "downscaled" version of the  $1^\circ$ -open-ocean data and the  $0.25^\circ$ -coastal-ocean data (see  
Sect. 2.1). Open-ocean SOCAT datasets gridded at finer regular resolutions (if accessible) will be updated to gain more accuracy  
in our model fitting. Selections of data products for predictors needed for model input are equally important. For instance, the  
CMEMS SSS product used here results in a globally good reconstruction of total alkalinity (Table 5). However, the temporal  
variability in CMEMS SSS data does not match that in observations (Fig. A10) and this feature is retained in time series of total  
650 alkalinity (Fig. 7). Despite best efforts in determining overall product uncertainty in estimates of carbonate system variables,  
part of input uncertainty is still not taken into account or only partially quantified due to lack of time-space varying uncertainty  
fields associated with predictor variables (e.g. SSH, Chl-a, MLD, nutrient concentrations). Moreover, temporal sampling bias  
in  $p\text{CO}_2$  and  $p\text{H}$  is likely to contribute to deviations between observations and model output (Fig. 5 and Table A3). The total  
measurement error uncertainty should be considered with great care during reconstruction and model output evaluation.

655 The CMEMS-LSCE approach leads as the first series of long-term reconstructions of  $p\text{CO}_2$ ,  $p\text{H}$ ,  $A_T$ , DIC,  $\Omega_{ca}$  and  $\Omega_{ar}$  ex-  
tending seamlessly from the global open ocean to coastal regions at monthly,  $0.25^\circ$  resolutions. Future use cases recommended  
for this high-resolution product include (1) estimation of monthly to interannual variations, long-term trends of carbonate sys-  
tem variables, as well as of air-sea  $\text{CO}_2$  exchanges at the surface layer from local scale to large ocean basins, (2) analyses  
in interactions between these variables and effects of other physical and biogeochemical factors on ocean acidification and  
660 changes in the marine carbonate system, (3) assessments of horizontal and temporal gradients of carbonate system variables  
in the coastal-open ocean continuum, (4) evaluation or combination with other model- or observation-based products (e.g.,  
Biogeochemistry Argo, Southern Ocean Carbon and Climate Observations and Modeling), and (5) improvements in coastal  
reconstructions based on observation system simulation experiments (e.g., with finer spatio-temporal model resolutions). The  
CMEMS-FFNN surface ocean carbon product at monthly,  $0.25^\circ$  resolutions will be accessible through the CMEMS data portal  
665 (see Sect. 7).

## 7 Data availability

The CMEMS-LSCE datasets of six carbonate system variables have been delivered to the European Copernicus Marine En-  
vironment Monitoring Service (CMEMS, Product ID: MULTI\_OBS\_GLO\_BIO\_CARBON\_SURFACE\_REP\_015\_008, DOI:  
10.48670/moi-00047). Since November 2022, the product with monthly and  $1^\circ$  resolutions is available at the CMEMS portal  
670 (Chau et al., 2022a, b). The CMEMS-LSCE data product at monthly and  $0.25^\circ$  resolutions proposed in this study will replace  
its coarser resolution version in due course. For the time being, the high-resolution data product described in this manuscript  
can be accessed via repository under data DOI: 10.14768/a2f0891b-763a-49e9-af1b-78ed78b16982 (Chau et al., 2023).



## Appendix A: Definitions of ocean carbonate system variables

Chemical reactions of dissolved CO<sub>2</sub> in seawater follow a series of the following equilibria,



where (g) and (aq) stand for a gas or the species in an aqueous solution. CO<sub>2</sub>(aq) refers to the combination of aqueous CO<sub>2</sub> and its weak acid H<sub>2</sub>CO<sub>3</sub>. HCO<sub>3</sub><sup>-</sup> and CO<sub>3</sub><sup>2-</sup> are bicarbonate and carbonate ions.

680 Definitions of the essential variables involved in the carbonate system equilibria (A1) are on the list below (see in Dickson et al., 2007; Dickson, 2010; Gattuso and Hansson, 2011, for further details).

- i) Surface ocean *p*CO<sub>2</sub> is partial pressure of CO<sub>2</sub> in air which is in equilibrium with that in water sample. It is not the same as surface ocean fugacity of CO<sub>2</sub> (*f*CO<sub>2</sub>). *p*CO<sub>2</sub> can be converted from *f*CO<sub>2</sub> via

$$p\text{CO}_2 = f\text{CO}_2 \exp\left(-P \frac{B + 2\delta}{RT^*}\right). \quad (\text{A2})$$

685 where *P* is total atmospheric pressure at surface water, *T*<sup>\*</sup> is absolute temperature, *R* is the gas constant, and *B* and *δ* are cross-virial coefficients (Körtzinger, 1999).

- ii) Seawater *p*H is a negative logarithmic scale of total concentration of hydrogen ions (H<sup>+</sup>) in aqueous solution. Total H<sup>+</sup> is the sum of concentrations of free H<sup>+</sup> and HSO<sub>4</sub> ions. The *p*H scale typically ranges from 0 to 14. *p*H = 7 is the threshold specifying whether a water sample is in acidic (i.e., *p*H < 7) or basic (i.e., *p*H > 7) conditions.

690 iii) Total alkalinity (A<sub>T</sub>) measures the capacity of seawater against acidification. By definition, A<sub>T</sub> is total concentration of dissolved alkaline substances corresponding to the ability in H<sup>+</sup> attracting over H<sup>+</sup> releasing. The major contributions to alkalinity includes bicarbonate (HCO<sub>3</sub><sup>-</sup>), carbonate (CO<sub>3</sub><sup>2-</sup>), and hydroxide (OH<sup>-</sup>) ions. Total alkalinity can be approximated with

$$A_T = [\text{HCO}_3^-] + 2[\text{CO}_3^{2-}] + [\text{OH}^-] - [\text{H}^+]. \quad (\text{A3})$$

695 iv) Total dissolved inorganic carbon (DIC) is the sum in concentrations of the three primary aqueous species in seawater,

$$\text{DIC} = [\text{HCO}_3^-] + [\text{CO}_3^{2-}] + [\text{CO}_2(\text{aq})]. \quad (\text{A4})$$

v) Calcium carbonate saturation state (*Ω*) is defined as follows,

$$\Omega = \frac{[\text{Ca}^{2+}][\text{CO}_3^{2-}]}{K_{sp}}, \quad (\text{A5})$$



700 where  $[Ca^{2+}]$  is the concentration of dissolved calcium ions and  $K_{sp}$  is the solubility of calcium carbonate in seawater.  $CaCO_3$  has two principal minerals: aragonite and calcite. Aragonite, which is more soluble than calcite ( $\Omega_{ar} < \Omega_{ca}$ ), is produced by many marine shells and skeletons including corals, pteropods, clams, and mussels. A  $\Omega_{ar}$  value greater than 1, i.e., preferable conditions in shell formation, indicates supersaturated seawater with respect to aragonite, and vice versa.

705 vi) Revelle factor (RF) measures the buffer capacity for the carbonate system in seawater that decreases as  $pH$  increases. Revelle factor is expressed by the ratio between instantaneous changes of dissolved  $CO_2$   $\left(\frac{[\Delta CO_2(aq)]}{[CO_2(aq)]}\right)$  and of DIC  $\left(\frac{[\Delta DIC]}{[DIC]}\right)$  in seawater,

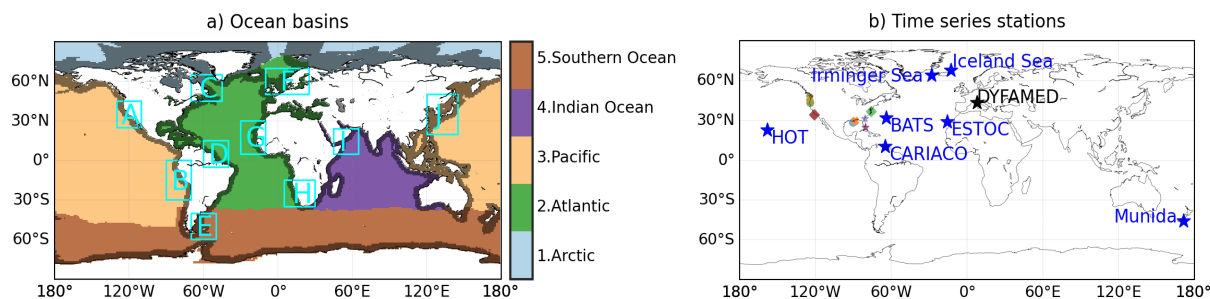
$$RF = \frac{[\Delta CO_2(aq)]}{[CO_2(aq)]} \left(\frac{[\Delta DIC]}{[DIC]}\right)^{-1}. \quad (A6)$$

710 *Author contributions.* TTTC, FC, and MG developed the CMEMS-LSCE-FFNN model at a quarter-degree resolution. TTTC has prepared script codes and executed the experiments with support from FC in setting and running the model at the HPC resources of TGCC. TTTC, MG, and FC shaped the first manuscript version. All the authors contribute to revise the manuscript based on NM's suggestions.

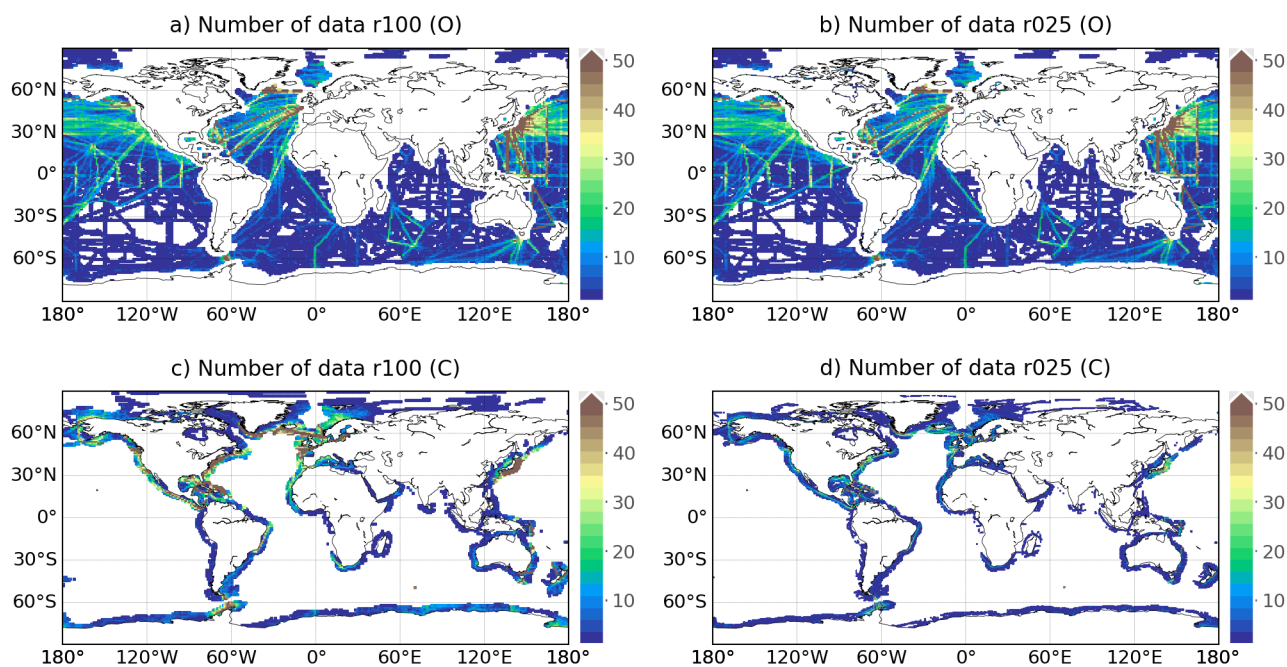
*Competing interests.* The author and co-authors have declared that they have no competing interests.

715 *Acknowledgements.* This research has been supported by the MOB TAC project of the European Copernicus Marine Environment Monitoring Service (CMEMS) (<https://marine.copernicus.eu/about/producers/mob-tac>, last access: 14 March 2023). It was granted access to the HPC resources of TGCC under the allocation A0110102201 made by GENCI. The Surface Ocean  $CO_2$  Atlas (SOCAT, [www.socat.info](http://www.socat.info), last access: 20 March 2023) is an international effort, endorsed by the International Ocean Carbon Coordination Project (IOCCP), the Surface Ocean Lower Atmosphere Study (SOLAS) and the Integrated Marine Biogeochemistry and Ecosystem Research program (IMBER), to deliver a uniformly quality-controlled surface ocean  $CO_2$  database. We thank Anna Conchon for her help in testing and wrapping LIAR and CO2SYS Matlab toolboxes.





**Figure A1.** a) Ocean basins (<https://github.com/RECCAP2-ocean/RECCAP2-shared-resources/tree/master/data/regions>, last access: 11/7/2022): coastal mask (grey, approximately 400 km from the shore line), feature regions analyzed in this study (cyan box, Table A1); b) Location of time series stations recording in situ observations used in data evaluation (Table 3): blue stars (Bates et al., 2014), black star (Coppola et al., 2021), and other coloured scattered objects (Sutton et al., 2019).



**Figure A2.** Spatial distribution of total months in 1985-2021 containing SOCATv2022 gridded data. Left:  $1^\circ$ -data product (r100), right:  $0.25^\circ$ -data product (r025). Open-ocean data (O) in each  $0.25^\circ$ -grid box is created by setting conservatively the open-ocean SOCATv2022 data at the  $1^\circ$ -grid box containing it. The coastal-ocean SOCATv2022 data (C) are assigned within 400 km from the shoreline (Fig. A1a).



**Table A1.** Information of feature regions analyzed in this study (Fig. A1a - cyan boxes).

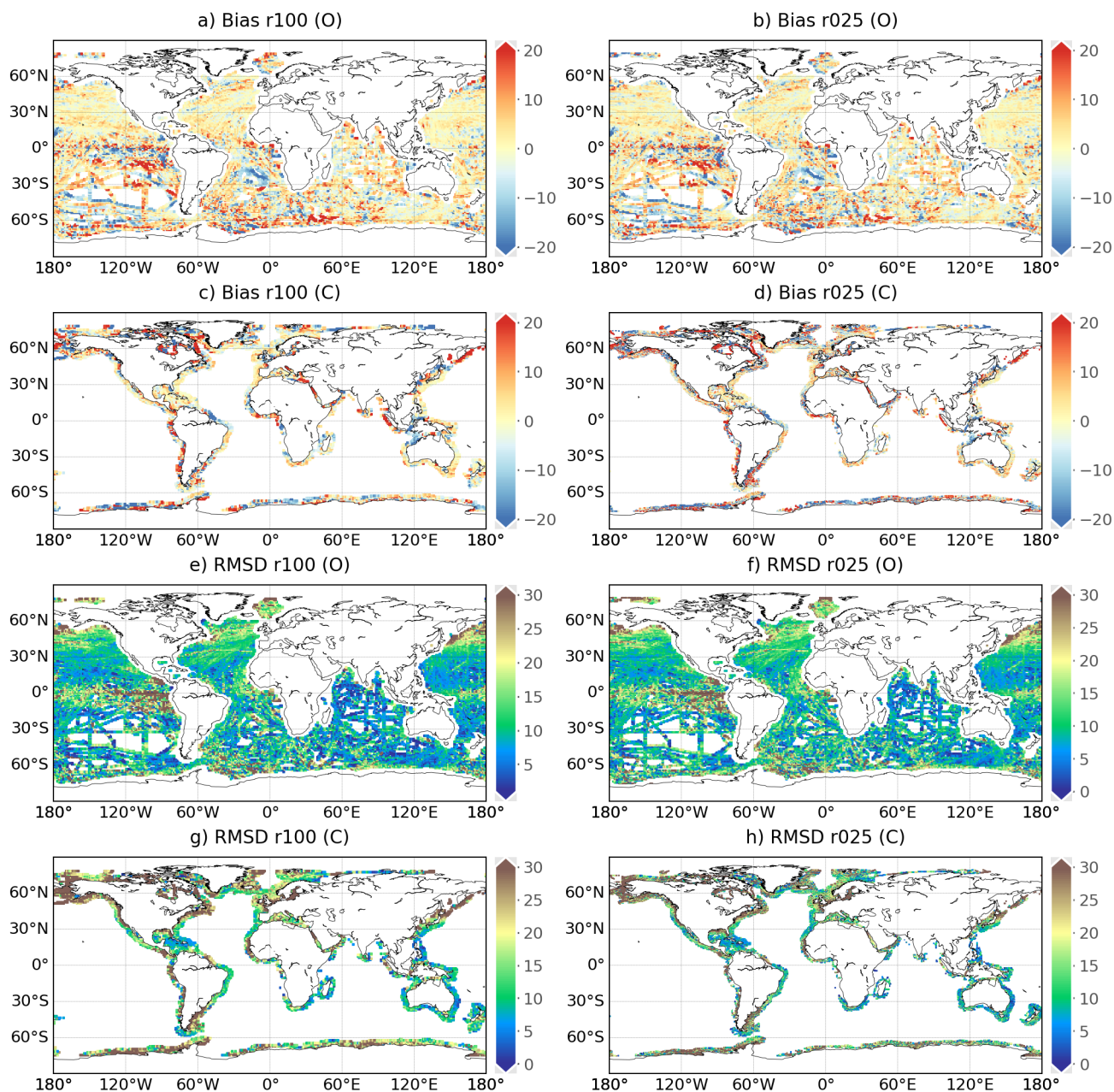
Notations	Regions	Coordinates	
		Latitude	Longitude
A	California Current System	25°N-45°N	130°W-110°W
B	Humboldt Current System	30°S-0°	90°W-70°W
C	Labrador Sea	45°N-65°N	70°W-45°W
D	Amazon river mouth	5°S-15°N	60°W-40°W
E	Western South Atlantic	60°S-40°S	70°W-50°W
F	Northern Europe	50°N-70°N	10°W-25°E
G	Canary Current System	5°N-30°N	30°W-10°W
H	Benguela Current System	35°S-15°S	5°E-20°E
I	Western Arabian Sea	5°N-24°N	45°E-65°E
J	Sea of Japan	30°N-50°N	120°E-150°E

**Table A2.** Information of moored time series of coastal-surface-ocean  $p\text{CO}_2$  and  $p\text{H}$  observations (Sutton et al., 2019).

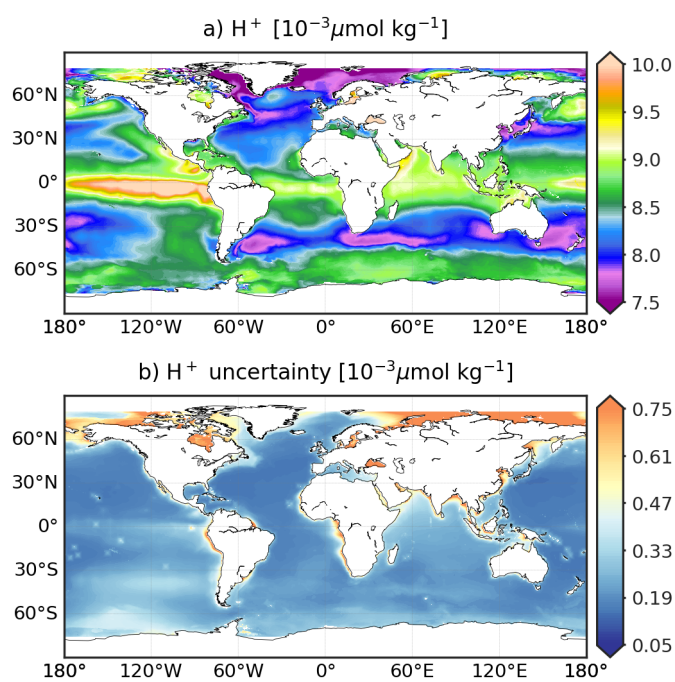
Stations	Abbreviations	Coordinates	Date range
1. Cape Arago	CAPEARAGO	43.3°N, 124.5°W	06/2017-12/2020
2. Cape Elizabeth	CAPEELIZABETH	47.4°N, 124.7°W	06/2006-05/2020
3. California Current Ecosystem 2	CCE2	34.3°N, 120.8°W	01/2010-06/2021
4. Cheeca Rocks Ocean Acidification Mooring in Florida Keys National Marine Sanctuary	CHEECAROCKS	24.9°N, 80.6°W	12/2011-12/2021
5. Coastal Louisiana buoy	COASTALLA	28.5°N, 90.3°W	07/2017-08/2020
6. Central Gulf of Mexico Ocean Observing System Station 01	COASTALMS	30.0°N, 88.6°W	05/2009-05/2017
7. Chesapeake Bay Interpretive Buoy System Ocean Acidification Buoy at First Landing	FIRSTLANDING	37.0°N, 76.1°W	04/2018-09/2020
8. NDBC Buoy 41008 in Gray's Reef National Marine Sanctuary	GRAYSREEF	31.4°N, 80.9°W	07/2006-08/2018

**Table A3.** Statistics computed between CMEMS-LSCE datasets (0.25°) and time series of  $p\text{CO}_2$  and  $p\text{H}$  measurements (Sutton et al., 2019): total numbers of monthly mean observations (N), temporal standard deviation of observations from their monthly averages ( $\sigma^t$ ), RMSD (Eq. 10), and  $r^2$  (Eq. 11). See Table A2 and Fig. A1b for stations' information and locations.

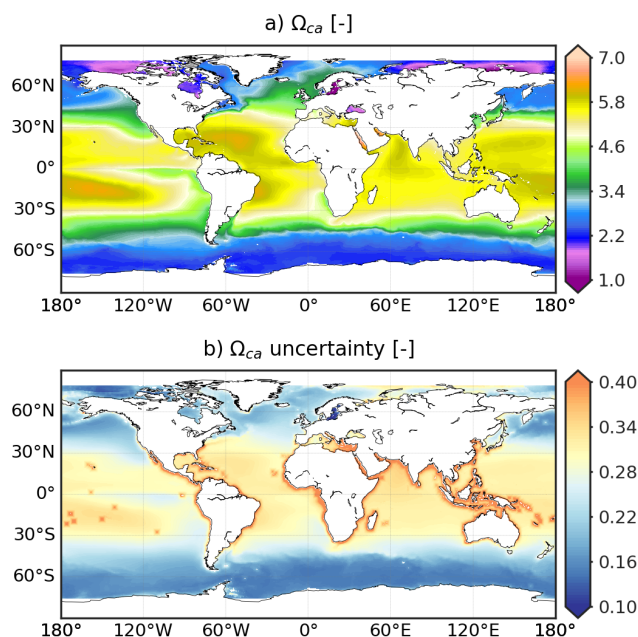
Stations	$p\text{CO}_2$ [ $\mu\text{atm}$ ]				$p\text{H}$ [-]			
	N	$\sigma^t$	RMSD	$r^2$	N	$\sigma^t$	RMSD	$r^2$
1. CAPEARAGO	33	65.60	79.86	0.19	31	0.086	0.069	0.22
2. CAPEELIZABETH	92	42.54	41.44	0.52	11	0.061	0.057	0.69
3. CCE2	127	45.31	32.44	0.16	58	0.048	0.035	0.24
4. CHEECAROCKS	73	44.05	62.42	0.25	40	0.038	0.066	0.21
5. COASTALLA	22	59.15	57.41	0.52	2	0.078	0.068	-
6. COASTALMS	41	43.04	42.50	0.51	15	0.062	0.065	0.25
7. FIRSTLANDING	17	69.98	77.32	0.49	2	0.061	0.042	-
8. GRAYSREEF	96	20.12	38.34	0.65	49	0.020	0.040	0.66



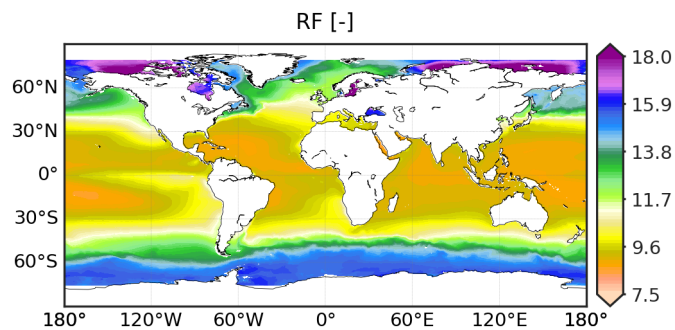
**Figure A3.** Global maps of mean model-data difference (Bias, abcd) and root-of-mean-square-deviation (RMSD, efgh) between the reconstruction and SOCATv2022  $p\text{CO}_2$  [ $\mu\text{atm}$ ] over 1985-2021. Left: CMEM-LSCE-FFNN with a resolution of  $1^\circ$  (r100), right: CMEM-LSCE-FFNN with a resolution of  $0.25^\circ$  (r025). Open-ocean data (O) in each  $0.25^\circ$ -grid box used for evaluation is created by setting conservatively the open-ocean SOCATv2022 data value at the  $1^\circ$ -grid box containing it. Coastal-ocean data (C) are extracted from each of the two SOCATv2022 gridded data products.



**Figure A4.** CMEMS-LSCE  $H^+$  over the global ocean at a spatial resolution of  $0.25^\circ$ . Temporal means of the model best estimate and  $1\sigma$ -uncertainty per grid cell over 1985-2021 are calculated by using Eq. (5).

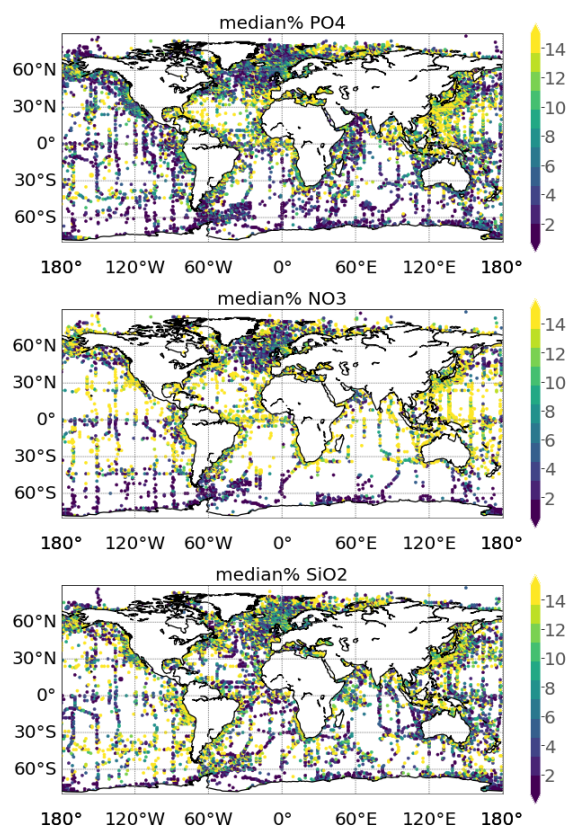


**Figure A5.** CMEMS-LSCE  $\Omega_{ca}$  over the global ocean at a spatial resolution of  $0.25^\circ$ . Temporal means of the model best estimate and  $1\sigma$ -uncertainty per grid cell over 1985-2021 are calculated by using Eq. (5).

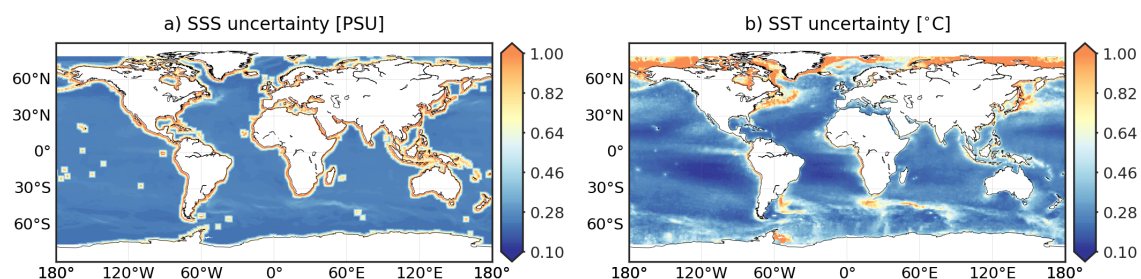


**Figure A6.** CMEMS-LSCE Revelle Factor (RF) over the global ocean at a spatial resolution of  $0.25^\circ$ . Temporal means of the model best estimate per grid cell over 1985-2021 are calculated by using Eq. (5).



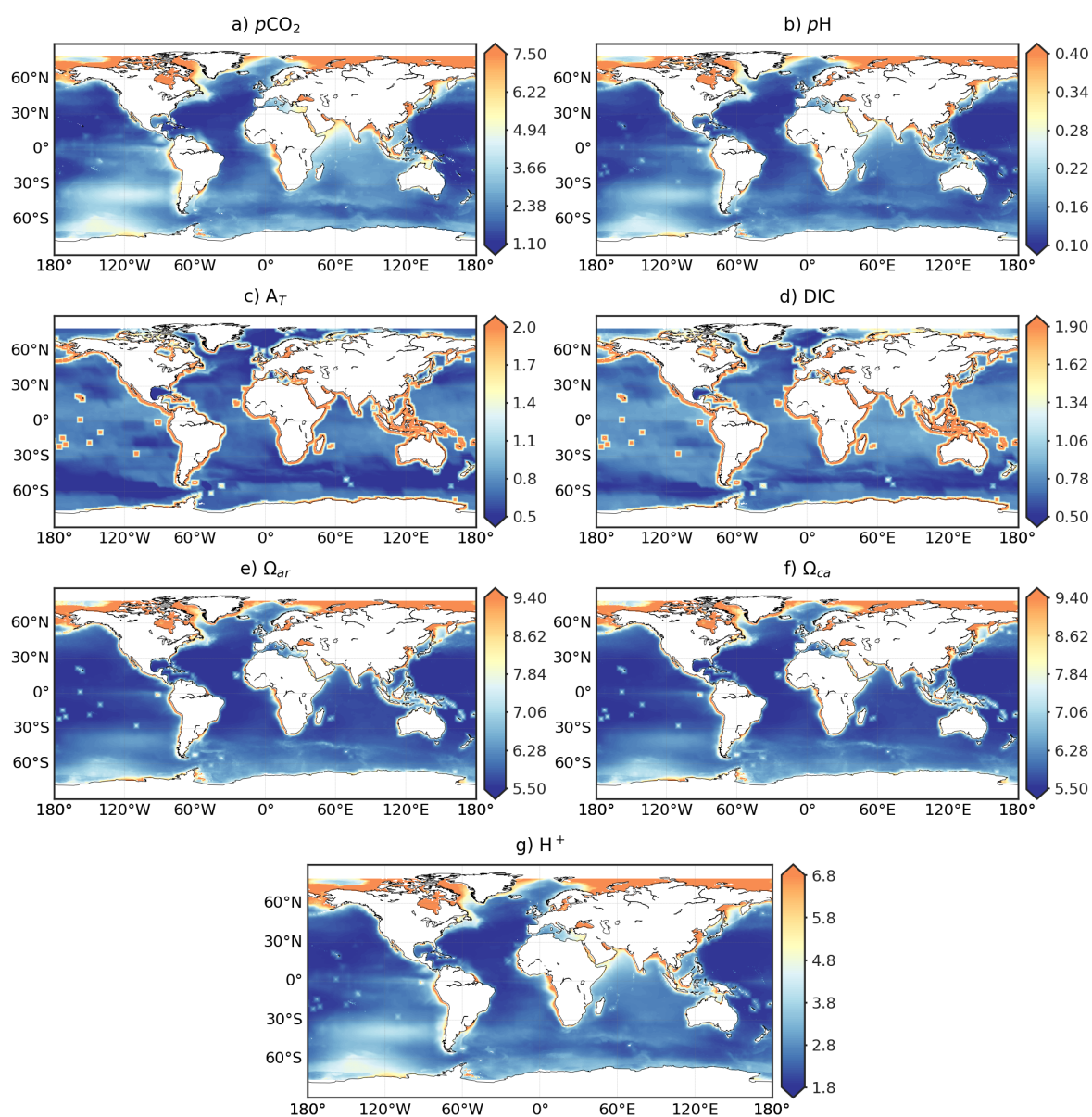


**Figure A7.** Median percentage of analysis error uncertainty against of climatological mean of surface WOA18 nutrient data: phosphate ( $\text{PO}_4$ ), nitrate ( $\text{NO}_3$ ), and silicate ( $\text{SiO}_2$ ).

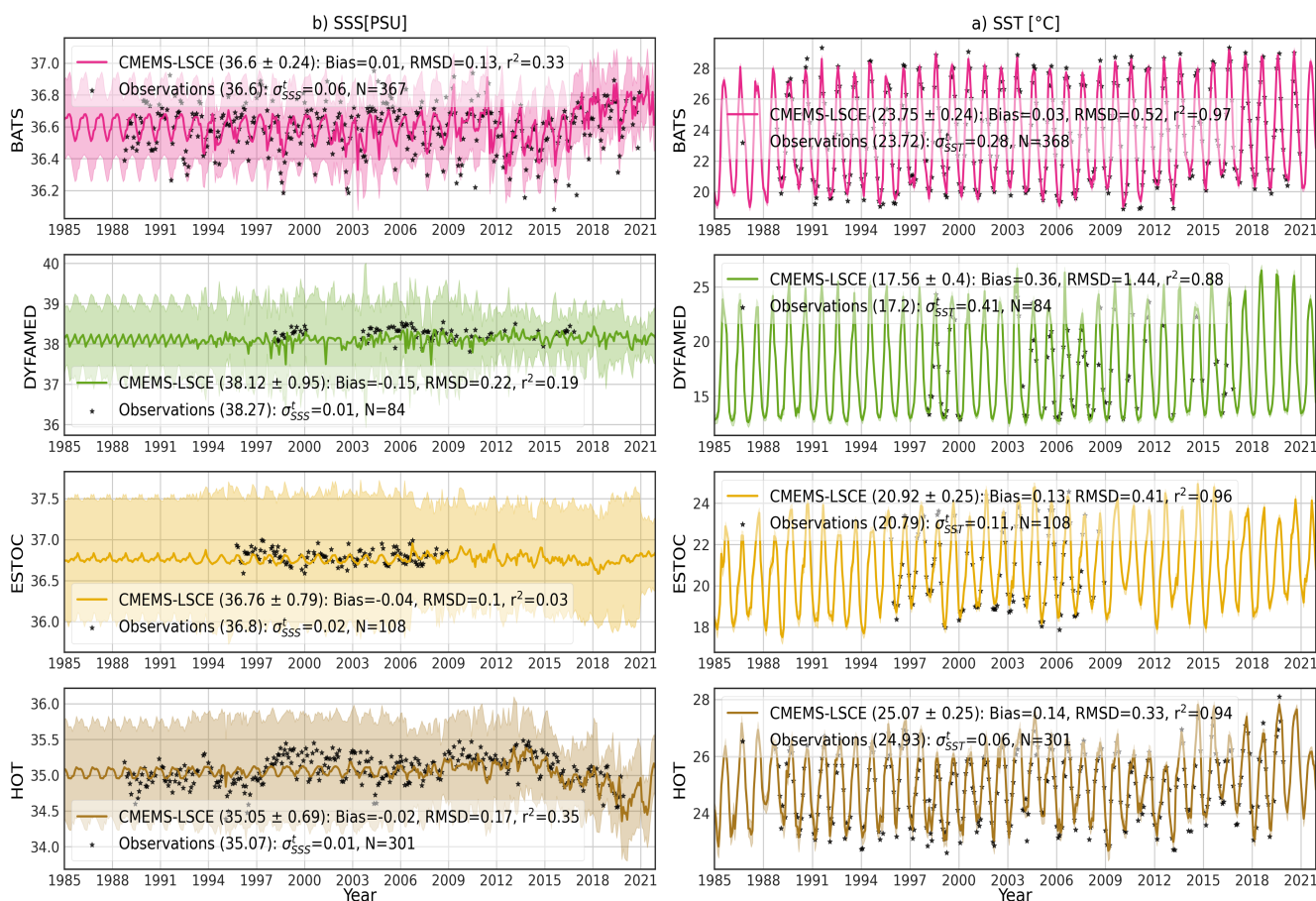


**Figure A8.** Spatial distribution of CMEMS SSS and SST product uncertainty over the global ocean at a spatial resolution of  $0.25^\circ$ .  $1\sigma$ -uncertainty is computed per grid cell by using Eq. (5) over 1985-2021.

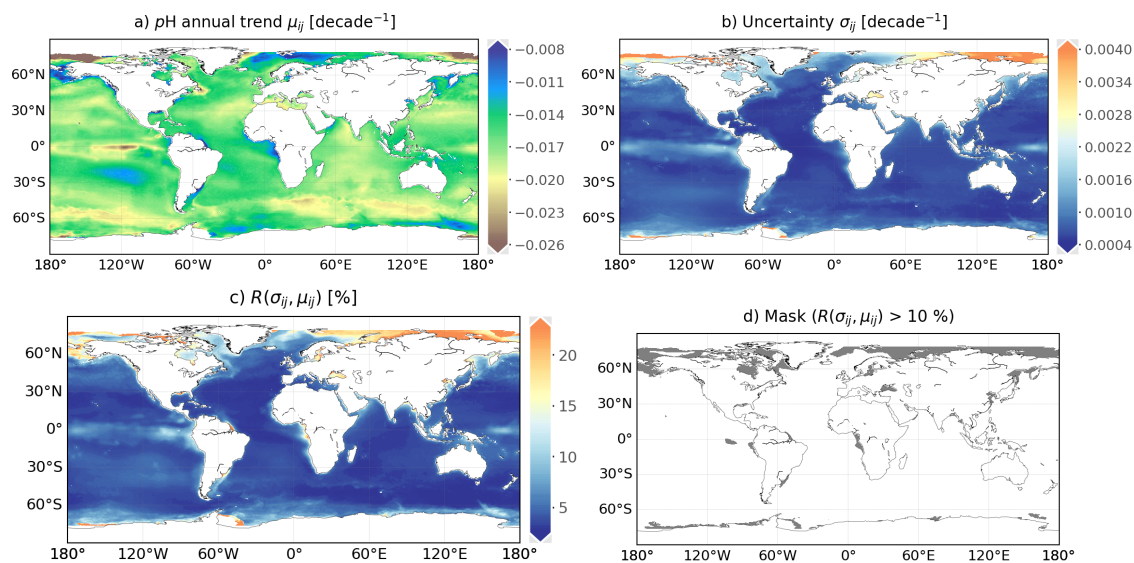




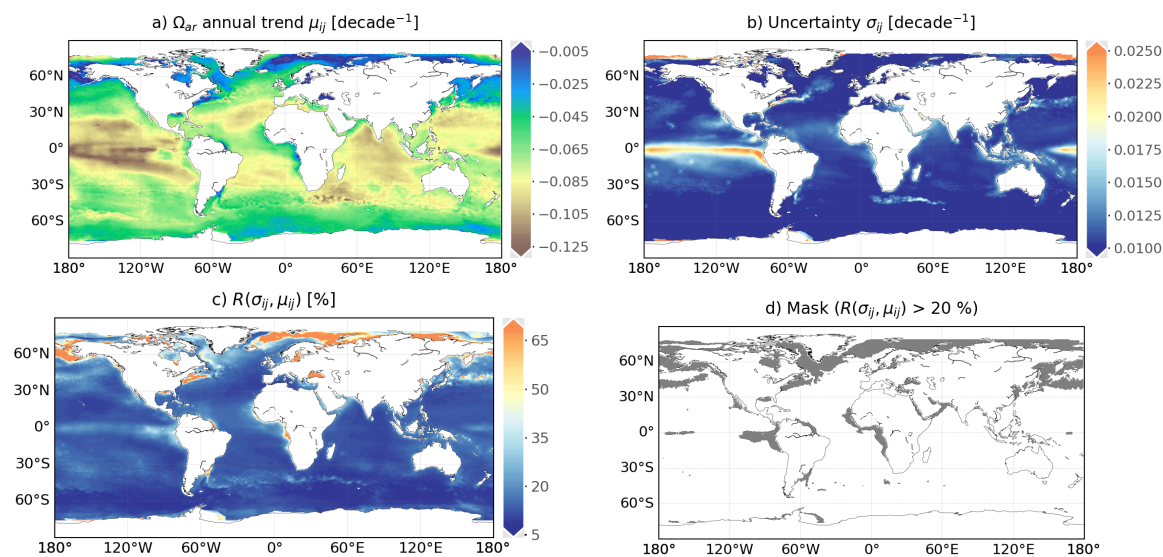
**Figure A9.** Spatial distribution of  $R(\sigma, \mu)$  [%] (Eq 8), i.e., the ratio of model uncertainty ( $\sigma$ ) against model best estimate ( $\mu$ ).



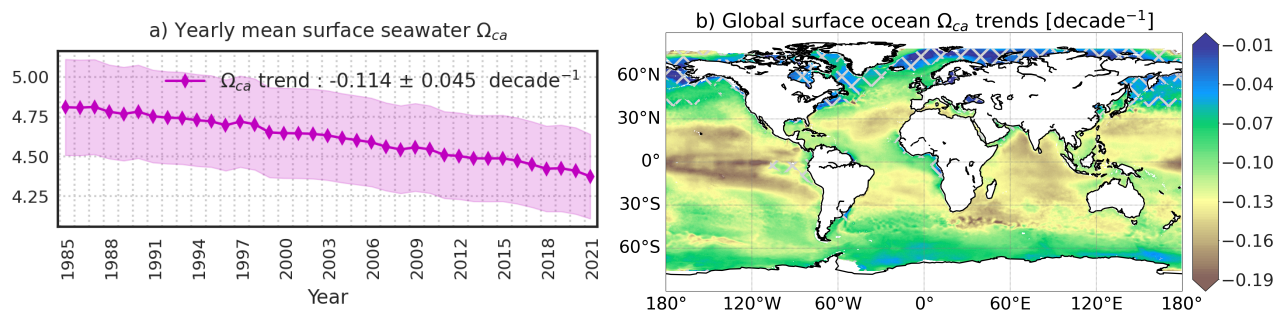
**Figure A10.** Monthly time series of SSS and SST at BATS, DYFAMED, ESTOC, and HOT stations (Table 3 and Fig. A1b): model best estimate (curve),  $1\sigma$ -uncertainty (envelope), and monthly average of observations (point). Means of the best estimate and  $1\sigma$ -uncertainty ( $\mu \pm \sigma$ ) calculated over the observing time span are shown in brackets if accessible. Statistics include number of months with observations ( $N$ ), Bias, RMSD, and  $r^2$ .  $\sigma_{SS}^t$  [ $\sigma_{SST}^t$ ] stands for temporal standard deviation from monthly averages of SSS and SST observations. Temporal variations in SSS observations are poorly described in the CMEMS SSS time series (Table 2) used in CMEMS-LSCE reconstructions.



**Figure A11.** a) Global surface seawater  $\rho\text{H}$  trend over the period 1985–2021, b)  $1\sigma$ -uncertainties associated to trend estimates, c)  $\sigma$ -to- $\mu$  ratio  $R(\sigma, \mu)$  [%] (Eq. 8) between uncertainty estimates (b) and the best trend estimates (a), d) mask applied over the regions where  $R(\mu, \sigma) > 10\%$ .

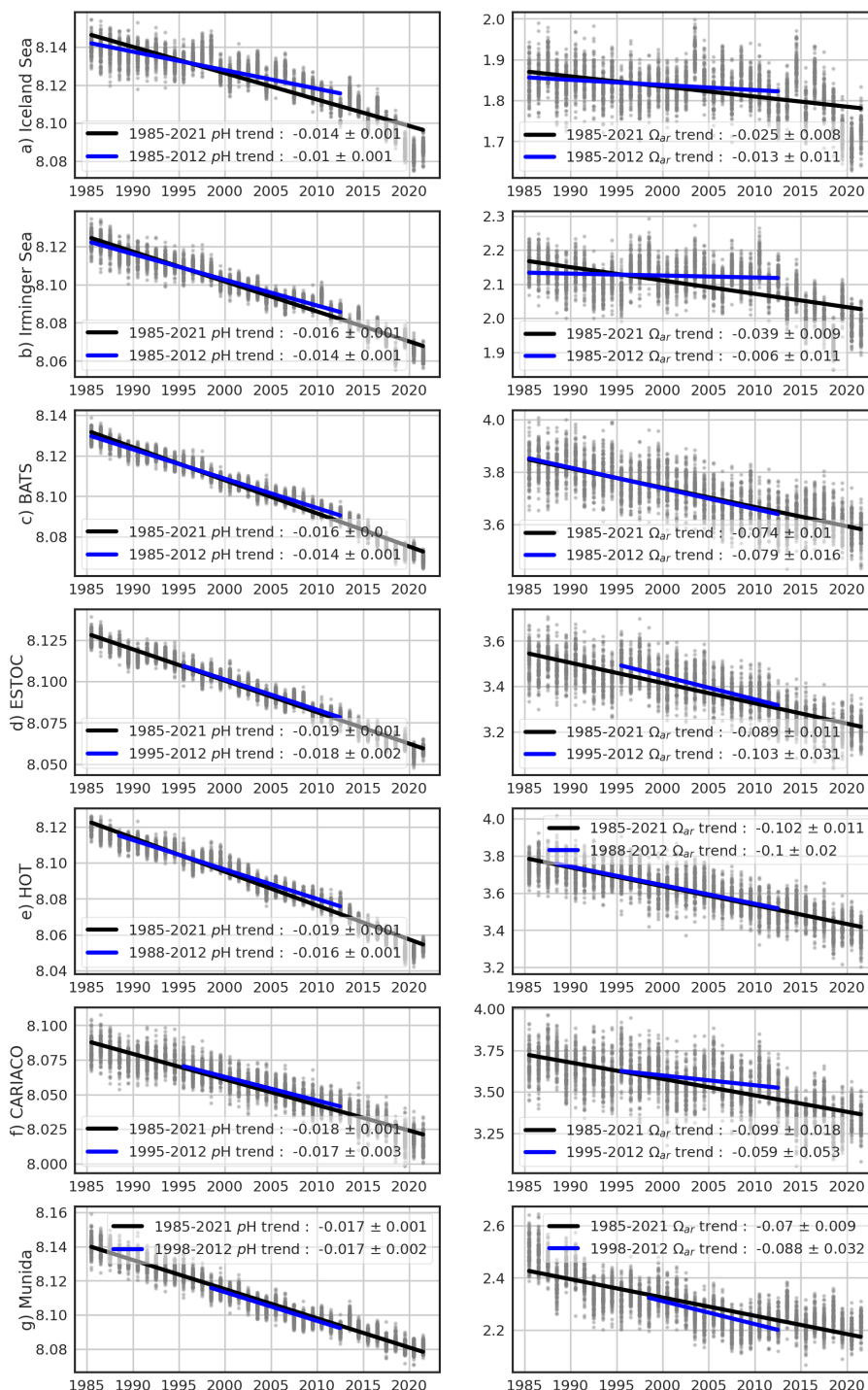


**Figure A12.** a) Global surface seawater  $\Omega_{ar}$  trend over the period 1985–2021, b)  $1\sigma$ -uncertainties associated to trend estimates, c)  $\sigma$ -to- $\mu$  ratio  $R(\sigma, \mu)$  [%] (Eq. 8) between uncertainty estimates (b) and the best trend estimates (a), d) mask applied over the regions where  $R(\mu, \sigma) > 20\%$ .



**Figure A13.** a) Yearly global area-weighted mean of surface ocean saturation states with respect to calcite ( $\Omega_{ca}$ ): Global means of the best estimate ( $\mu$ , plain line) and uncertainty ( $\sigma$ , envelop) are computed with Eq. (7a). b) Global trend maps of  $\Omega_{ca}$  over 1985-2021: Cross-hatching covers the regions where uncertainty of a trend estimate over 20% of the trend value.





**Figure A14.** Linear trend estimates learned on 100-member ensemble (grey points) of yearly mean time series of  $pH$  and  $\Omega_{ar}$  at different stations (Bates et al., 2014).  $\mu \pm \sigma$  present linear slope and residual standard deviation. Black or blue lines stand for linear fits over the full or sub-period in 1985-2021 (see Table 6 for comparison).





## References

- 720 Bakker, D., Alin, S., Becker, M., Bittig, H., Castaño-Primo, R., Feely, R. A., Gritzalis, T., Kadono, K., Kozyr, A., Lauvset, S. K., Metzl, N., Munro, D., Nakaoka, S.-i., Nojiri, Y., O'Brien, K., Olsen, A., Pfeil, B., Pierrot, D., Steinhoff, T., Sullivan, K., Sutton, A., Sweeney, C., Tilbrook, B., Wada, C., Wanninkhof, R., Wranne, A. W., et al.: SOCAT version 2022 for quantification of ocean CO<sub>2</sub> uptake, [https://www.socat.info/wp-content/uploads/2022/06/2022\\_Poster\\_SOCATv2022\\_release.pdf](https://www.socat.info/wp-content/uploads/2022/06/2022_Poster_SOCATv2022_release.pdf), 2022.
- 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.,  
725 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., Ibáñez, 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.,  
730 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 *f*CO<sub>2</sub> data in version 3 of the Surface Ocean CO<sub>2</sub> Atlas (SOCAT), *Earth System Science Data*, 8, 383–413, <https://doi.org/10.5194/essd-8-383-2016>, 2016.
- 735 Bates, N. R., Astor, Y. M., Church, M. J., Currie, K., Dore, J. E., González-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J., and Santana-Casiano, J. M.: A time-series view of changing surface ocean chemistry due to ocean uptake of anthropogenic CO<sub>2</sub> and ocean acidification, *Oceanography*, 27, 126–141, 2014.
- Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and  
740 Vichi, M.: Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models, *Biogeosciences*, 10, 6225–6245, <https://doi.org/10.5194/bg-10-6225-2013>, 2013.
- Bresnahan, P. J., Takeshita, Y., Wirth, T., Martz, T. R., Cyronak, T., Albright, R., Wolfe, K., Warren, J. K., and Mertz, K.: Autonomous in situ calibration of ion-sensitive field effect transistor pH sensors, *Limnology and Oceanography: Methods*, 19, 132–144, 2021.
- Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., Key, R. M., Tanhua, T., González-Dávila, M., Jeansson, E., et al.: A global monthly climatology of total alkalinity: a neural network approach, *Earth System Science Data*, 11, 1109–1127, 2019.
- 745 Canadell, J. G., Monteiro, P. M., Costa, M. H., Da Cunha, L. C., Cox, P. M., Alexey, V., Henson, S., Ishii, M., Jaccard, S., Koven, C., et al.: Global carbon and other biogeochemical cycles and feedbacks In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, <https://doi.org/10.1017/9781009157896.007>, 2021.
- Carter, B. R., Williams, N. L., Gray, A. R., and Feely, R. A.: Locally interpolated alkalinity regression for global alkalinity estimation,  
750 *Limnology and Oceanography: Methods*, 14, 268–277, <https://doi.org/10.1002/lom3.10087>, 2016.
- Carter, B. R., Feely, R. A., Williams, N. L., Dickson, A. G., Fong, M. B., and Takeshita, Y.: Updated methods for global locally interpolated estimation of alkalinity, pH, and nitrate, *Limnology and Oceanography: Methods*, 16, 119–131, <https://doi.org/10.1002/lom3.10232>, 2018.
- Chau, T. T. T., Gehlen, M., and Chevallier, F.: Global ocean surface carbon product, Research Report CMEMS-MOB-QUID-015-008, Le  
755 Laboratoire des Sciences du Climat et de l'Environnement, <https://doi.org/10.48670/moi-00047>, quality Information Document, 2022a.



- Chau, T. T. T., Gehlen, M., and Chevallier, F.: A seamless ensemble-based reconstruction of surface ocean  $p\text{CO}_2$  and air–sea  $\text{CO}_2$  fluxes over the global coastal and open oceans, *Biogeosciences*, 19, 1087–1109, <https://doi.org/10.5194/bg-19-1087-2022>, 2022b.
- Chau, T. T. T., Gehlen, M., Metzl, N., and Chevallier, F.: CMEMS-LSCE: A global 0.25-degree, monthly reconstruction of the surface ocean carbonate system, ESPRI/IPSL [dataset], <https://doi.org/10.14768/a2f0891b-763a-49e9-af1b-78ed78b16982>, 2023.
- 760 Chevallier, F.: On the parallelization of atmospheric inversions of  $\text{CO}_2$  surface fluxes within a variational framework, *Geoscientific Model Development*, 6, 783–790, <https://doi.org/10.5194/gmd-6-783-2013>, 2013.
- Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring  $\text{CO}_2$  sources and sinks from satellite observations: Method and application to TOVS data, *Journal of Geophysical Research: Atmospheres*, 110, <https://doi.org/https://doi.org/10.1029/2005JD006390>, 2005.
- 765 Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguá, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.:  $\text{CO}_2$  surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *Journal of Geophysical Research: Atmospheres*, 115, <https://doi.org/https://doi.org/10.1029/2010JD013887>, 2010.
- 770 Cooley, S., Schoeman, D., Bopp, L., Boyd, P., Donner, S., Ito, S.-i., Kiessling, W., Martinetto, P., Ojea, E., Racault, M.-F., et al.: Oceans and coastal ecosystems and their services, in: IPCC AR6 WGII, Cambridge University Press, 2022.
- Coppola, L., Boutin, J., Gattuso, J.-P., Lefèvre, D., and Metzl, N.: The carbonate system in the Ligurian Sea, The Mediterranean Sea in the era of global change 1: 30 years of multidisciplinary study of the ligurian Sea, pp. 79–103, 2020.
- Coppola, L., Diamond, R. E., and Carval, T.: Dyfamed observatory data. SEANOE, <https://doi.org/0.17882/43749>, 2021.
- 775 Denvil-Sommer, A., Gehlen, M., Vrac, M., and Mejia, C.: LSCE-FFNN-v1: a two-step neural network model for the reconstruction of surface ocean  $p\text{CO}_2$  over the global ocean, *Geoscientific Model Development*, 12, 2091–2105, <https://doi.org/10.5194/gmd-12-2091-2019>, 2019.
- Dickson, A. G.: Standard potential of the reaction  $-\text{AgCl}(\text{s}) + 1/2\text{H}_2(\text{g}) = \text{Ag}(\text{s}) + \text{HCl}(\text{aq})$  and the standard acidity constant of the ion  $\text{HSO}_4^-$  in synthetic sea-water from 273.15-K to 318.15-K, *Journal of Chemical Thermodynamics*, 22, 113–127, [https://doi.org/doi:10.1016/0021-9614\(90\)90074-z](https://doi.org/doi:10.1016/0021-9614(90)90074-z), 1990.
- 780 Dickson, A. G.: The carbon dioxide system in seawater: equilibrium chemistry and measurements in Guide to best practices for ocean acidification research and data reporting, [https://www.pmel.noaa.gov/co2/files/dickson\\_thecarbondioxidesysteminseawater\\_equilibriumchemistryandmeasurementspp17-40.pdf](https://www.pmel.noaa.gov/co2/files/dickson_thecarbondioxidesysteminseawater_equilibriumchemistryandmeasurementspp17-40.pdf), 2010.
- Dickson, A. G., Sabine, C. L., and Christian, J. R.: Guide to best practices for ocean  $\text{CO}_2$  measurements., North Pacific Marine Science Organization, 2007.
- 785 Doney, S. C.: The growing human footprint on coastal and open-ocean biogeochemistry, *science*, 328, 1512–1516, 2010.
- Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.: Ocean acidification: the other  $\text{CO}_2$  problem, *Annual review of marine science*, 1, 169–192, 2009.
- Doney, S. C., Busch, D. S., Cooley, S. R., and Kroeker, K. J.: The impacts of ocean acidification on marine ecosystems and reliant human communities, *Annual Review of Environment and Resources*, 45, 2020.
- 790 Dore, J. E., Lukas, R., Sadler, D. W., Church, M. J., and Karl, D. M.: Physical and biogeochemical modulation of ocean acidification in the central North Pacific, *Proceedings of the National Academy of Sciences*, 106, 12 235–12 240, 2009.
- Droghei, R., Buongiorno Nardelli, B., and Santoleri, R.: A new global sea surface salinity and density dataset from multivariate observations (1993–2016), *Frontiers in Marine Science*, 5, 84, 2018.



- 795 Fabry, V. J., Seibel, B. A., Feely, R. A., and Orr, J. C.: Impacts of ocean acidification on marine fauna and ecosystem processes, *ICES Journal of Marine Science*, 65, 414–432, 2008.
- Feely, R. A., Wanninkhof, R., Takahashi, T., and Tans, P.: Influence of El Niño on the equatorial Pacific contribution to atmospheric CO<sub>2</sub> accumulation, *Nature*, 398, 597–601, 1999.
- Feely, R. A., Sabine, C. L., Takahashi, T., and Wanninkhof, R.: Uptake And Storage Of Carbon Dioxide In The Ocean: The Global CO<sub>2</sub> Survey, *OCEANOGRAPHY-WASHINGTON DC-OCEANOGRAPHY SOCIETY-*, 14, 18–32, 2001.
- 800 Friedlingstein, P., O’Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijckx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., Arneeth, A., Arora, V. K., Bates, N. R., Becker, M., Bellouin, N., Bittig, H. C., Bopp, L., Chevallier, F., Chini, L. P., Cronin, M., Evans, W., Falk, S., Feely, R. A., Gasser, T., Gehlen, M., Gkritzalis, T., Gloege, L., Grassi, G., Gruber, N., Gürses, O., Harris, I., Hefner, M., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jain, A. K., Jersild, A., Kadono, K., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I.,
- 805 Landschützer, P., Lefèvre, N., Lindsay, K., Liu, J., Liu, Z., Marland, G., Mayot, N., McGrath, M. J., Metz, N., Monacci, N. M., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O’Brien, K., Ono, T., Palmer, P. I., Pan, N., Pierrot, D., Pocock, K., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Rodriguez, C., Rosan, T. M., Schwinger, J., Séférian, R., Shutler, J. D., Skjelvan, I., Steinhoff, T., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tanhua, T., Tans, P. P., Tian, X., Tian, H., Tilbrook, B., Tsjino, H., Tubiello, F., van der Werf, G. R., Walker, A. P., Wanninkhof, R., Whitehead, C., Willstrand Wranne, A., Wright, R., Yuan, W., Yue, C., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.:
- 810 Global Carbon Budget 2022, *Earth System Science Data*, 14, 4811–4900, <https://doi.org/10.5194/essd-14-4811-2022>, 2022.
- Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, M., Zweng, M., Mishonov, A., Baranova, O., Seidov, D., et al.: World ocean atlas 2018. Vol. 4: Dissolved inorganic nutrients (phosphate, nitrate and nitrate+ nitrite, silicate), <https://archimer.ifremer.fr/doc/00651/76336/>, 2019.
- Gattuso, J.-P. and Hansson, L.: *Ocean acidification*, Oxford university press, 2011.
- 815 Gattuso, J.-P., Magnan, A., Billé, R., Cheung, W. W. L., Howes, E. L., Joos, F., Allemand, D., Bopp, L., Cooley, S. R., Eakin, C. M., Hoegh-Guldberg, O., Kelly, R. P., Pörtner, H.-O., Rogers, A. D., Baxter, J. M., Laffoley, D., Osborn, D., Rankovic, A., Rochette, J., Sumaila, U. R., Treyer, S., and Turley, C.: Contrasting futures for ocean and society from different anthropogenic CO<sub>2</sub> emissions scenarios, *Science*, 349, aac4722, <https://doi.org/10.1126/science.aac4722>, 2015.
- Gehlen, M., Gruber, N., Gangstø, R., Bopp, L., and Oschlies, A.: Biogeochemical consequences of ocean acidification and feedbacks to the
- 820 earth system, *Ocean acidification*, 1, 230–248, 2011.
- González-Dávila, M. and Santana-Casiano, J.: Sea surface and atmospheric fco2 data measured during the estoc time series cruises from 1995-2009, CDIAC, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee. doi, 10, 2009.
- Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J., and Worsfold, M.: The Current Configuration of the OSTIA System for Operational Production of Foundation Sea Surface Temperature and Ice Concentration
- 825 Analyses, *Remote Sensing*, 12, <https://www.mdpi.com/2072-4292/12/4/720>, 2020.
- Gregor, L. and Gruber, N.: OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification, *Earth System Science Data*, 13, 777–808, <https://doi.org/10.5194/essd-13-777-2021>, 2021.
- Gregor, L., Lebehot, A. D., Kok, S., and Scheel Monteiro, P. M.: A comparative assessment of the uncertainties of global surface ocean CO<sub>2</sub> estimates using a machine-learning ensemble (CSIR-ML6 version 2019a) – have we hit the wall?, *Geoscientific Model Development*, 12,
- 830 5113–5136, <https://doi.org/10.5194/gmd-12-5113-2019>, 2019.



- Gruber, N., Clement, D., Carter, B. R., Feely, R. A., Van Heuven, S., Hoppema, M., Ishii, M., Key, R. M., Kozyr, A., Lauvset, S. K., et al.:  
The oceanic sink for anthropogenic CO<sub>2</sub> from 1994 to 2007, *Science*, 363, 1193–1199, 2019.
- Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C., Bopp, L., Chau, T. T. T., Gürses, Ö., Ilyina, T., Landschützer, P., Lenton,  
A., Resplandy, L., Rödenbeck, C., Schwinger, J., and Séférian, R.: Consistency and Challenges in the Ocean Carbon Sink Estimate for the  
835 Global Carbon Budget, *Frontiers in Marine Science*, 7, 852, <https://doi.org/10.3389/fmars.2020.571720>, 2020.
- Hoffmann, L. J., Breitbarth, E., Boyd, P. W., and Hunter, K. A.: Influence of ocean warming and acidification on trace metal biogeochemistry,  
*Marine Ecology Progress Series*, 470, 191–205, 2012.
- Hopkins, F. E., Suntharalingam, P., Gehlen, M., Andrews, O., Archer, S. D., Bopp, L., Buitenhuis, E., Dadou, I., Duce, R., Goris, N., et al.:  
840 The impacts of ocean acidification on marine trace gases and the implications for atmospheric chemistry and climate, *Proceedings of the  
Royal Society A*, 476, 20190769, 2020.
- Ibáñez, J. S. P., Diverrès, D., Araujo, M., and Lefèvre, N.: Seasonal and interannual variability of sea-air CO<sub>2</sub> fluxes in the tropical Atlantic  
affected by the Amazon River plume, *Global Biogeochemical Cycles*, 29, 1640–1655, 2015.
- Iida, Y., Takatani, Y., Kojima, A., and Ishii, M.: Global trends of ocean CO<sub>2</sub> sink and ocean acidification: an observation-based reconstruction  
of surface ocean inorganic carbon variables, *Journal of Oceanography*, 77, 323–358, 2021.
- 845 Jiang, L.-Q., Carter, B. R., Feely, R. A., Lauvset, S. K., and Olsen, A.: Surface ocean pH and buffer capacity: past, present and future,  
*Scientific reports*, 9, 1–11, 2019.
- Jiang, L.-Q., Dunne, J., Carter, B. R., Tjiputra, J. F., Terhaar, J., Sharp, J. D., Olsen, A., Alin, S., Bakker, D. C. E., Feely, R. A., Gat-  
tuso, J.-P., Hogan, P., Ilyina, T., Lange, N., Lauvset, S. K., Lewis, E. R., Lovato, T., Palmieri, J., Santana-Falcón, Y., Schwinger, J.,  
Séférian, R., Strand, G., Swart, N., Tanhua, T., Tsujino, H., Wanninkhof, R., Watanabe, M., Yamamoto, A., and Ziehn, T.: Global  
850 Surface Ocean Acidification Indicators From 1750 to 2100, *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003563,  
<https://doi.org/https://doi.org/10.1029/2022MS003563>, 2023.
- Jones, S., Le Quéré, C., and Rödenbeck, C.: Autocorrelation characteristics of surface ocean pCO<sub>2</sub> and air-sea CO<sub>2</sub> fluxes, *Global biogeo-  
chemical cycles*, 26, 2012.
- Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., Feely, R. A., Millero, F. J., Mordy, C., and Peng, T.-  
855 H.: A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP), *Global Biogeochemical Cycles*, 18,  
<https://doi.org/https://doi.org/10.1029/2004GB002247>, 2004.
- Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., Lenton,  
A., Li, H., Lovenduski, N. S., Orr, J. C., Palmieri, J., Santana-Falcón, Y., Schwinger, J., Séférian, R., Stock, C. A., Tagliabue, A., Takano,  
Y., Tjiputra, J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A., and Ziehn, T.: Twenty-first century ocean warming,  
860 acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, *Biogeosciences*,  
17, 3439–3470, <https://doi.org/10.5194/bg-17-3439-2020>, 2020.
- Körtzinger, A.: Determination of carbon dioxide partial pressure (pCO<sub>2</sub>), chap. 9, pp. 149–158, John Wiley & Sons, Ltd,  
<https://doi.org/https://doi.org/10.1002/9783527613984.ch9>, 1999.
- Landschützer, P., Gruber, N., and Bakker, D. C.: Decadal variations and trends of the global ocean carbon sink, *Global Biogeochemical  
865 Cycles*, 30, 1396–1417, 2016.
- Landschützer, P., Ilyina, T., and Lovenduski, N. S.: Detecting Regional Modes of Variability in Observation-Based Surface Ocean pCO<sub>2</sub>,  
*Geophysical Research Letters*, 46, 2670–2679, 2019.



- Landschützer, P., Laruelle, G. G., Roobaert, A., and Regnier, P.: A uniform  $p\text{CO}_2$  climatology combining open and coastal oceans, *Earth System Science Data*, 12, 2537–2553, <https://doi.org/10.5194/essd-12-2537-2020>, 2020.
- 870 Laruelle, G. G., Landschützer, P., Gruber, N., Tison, J.-L., Delille, B., and Regnier, P.: Global high-resolution monthly  $p\text{CO}_2$  climatology for the coastal ocean derived from neural network interpolation, *Biogeosciences*, 14, 4545–4561, 2017.
- Lauvset, S. K. and Gruber, N.: Long-term trends in surface ocean pH in the North Atlantic, *Marine Chemistry*, 162, 71–76, 2014.
- Lauvset, S. K., Gruber, N., Landschützer, P., Olsen, A., and Tjiputra, J.: Trends and drivers in global surface ocean pH over the past 3 decades, *Biogeosciences*, 12, 1285–1298, <https://doi.org/10.5194/bg-12-1285-2015>, 2015.
- 875 Lauvset, S. K., Lange, N., Tanhua, T., Bittig, H. C., Olsen, A., Kozyr, A., Alin, S., Álvarez, M., Azetsu-Scott, K., Barbero, L., Becker, S., Brown, P. J., Carter, B. R., da Cunha, L. C., Feely, R. A., Hoppema, M., Humphreys, M. P., Ishii, M., Jeansson, E., Jiang, L.-Q., Jones, S. D., Lo Monaco, C., Murata, A., Müller, J. D., Pérez, F. F., Pfeil, B., Schirnack, C., Steinfeldt, R., Suzuki, T., Tilbrook, B., Ulfsbo, A., Velo, A., Woosley, R. J., and Key, R. M.: GLODAPv2.2022: the latest version of the global interior ocean biogeochemical data product, *Earth System Science Data*, 14, 5543–5572, <https://doi.org/10.5194/essd-14-5543-2022>, 2022.
- 880 Lee, K., Tong, L. T., Millero, F. J., Sabine, C. L., Dickson, A. G., Goyet, C., Park, G.-H., Wanninkhof, R., Feely, R. A., and Key, R. M.: Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans, *Geophysical research letters*, 33, 2006.
- Leseurre, C., Lo Monaco, C., Reverdin, G., Metzl, N., Fin, J., Mignon, C., and Benito, L.: Summer trends and drivers of sea surface  $f\text{CO}_2$  and pH changes observed in the southern Indian Ocean over the last two decades (1998–2019), *Biogeosciences*, 19, 2599–2625, <https://doi.org/10.5194/bg-19-2599-2022>, 2022.
- 885 Lewis, E. and Wallace, D.: CO2SYS-Program developed for the CO2 system calculations, Carbon Dioxide Inf Anal Center Report ORNL/CDIAC-105, <https://doi.org/10.15485/1464255>, 1998.
- Lueker, T. J., Dickson, A. G., and Keeling, C. D.: Ocean  $p\text{CO}_2$  calculated from dissolved inorganic carbon, alkalinity, and equations for K1 and K2: validation based on laboratory measurements of  $\text{CO}_2$  in gas and seawater at equilibrium, *Marine chemistry*, 70, 105–119, 2000.
- 890 Ma, D., Gregor, L., and Gruber, N.: Observation-based trends and drivers of global surface ocean acidification over the past four decades, *ESS Open Archive*, <https://doi.org/10.22541/essoar.167870335.58952668/v1>, 2023.
- Maritorena, S., d'Andon, O. H. F., Mangin, A., and Siegel, D. A.: Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues, *Remote Sensing of Environment*, 114, 1791–1804, <https://doi.org/10.1016/j.rse.2010.04.002>, 2010.
- 895 Menemenlis, D., Campin, J., Heimbach, P., Hill, C., Lee, T., Nguyen, A., Schodlok, M., and Zhang, H.: ECCO2: High Resolution Global Ocean and Sea Ice Data Synthesis, 2008, OS31C-1292, 2008.
- Menviel, L., Spence, P., Yu, J., Chamberlain, M., Matear, R., Meissner, K., and England, M. H.: Southern Hemisphere westerlies as a driver of the early deglacial atmospheric  $\text{CO}_2$  rise, *Nature Communications*, 9, 1–12, 2018.
- Michaels, A. F. and Knap, A. H.: Overview of the US JGOFS Bermuda Atlantic Time-series Study and the Hydrostation S program, *Deep Sea Research Part II: Topical Studies in Oceanography*, 43, 157–198, 1996.
- 900 Millero, F. J., Woosley, R., Ditrolio, B., and Waters, J.: Effect of ocean acidification on the speciation of metals in seawater, *Oceanography*, 22, 72–85, 2009.
- Mucci, A.: The solubility of calcite and aragonite in seawater at various salinities, temperatures, and one atmosphere total pressure, *American Journal of Science*, 283, 780–799, 1983.



- 905 Nardelli, B. B., Droghei, R., and Santoleri, R.: Multi-dimensional interpolation of SMOS sea surface salinity with surface temperature and in situ salinity data, *Remote Sensing of Environment*, 180, 392–402, 2016.
- Olivier, L., Boutin, J., Reverdin, G., Lefèvre, N., Landschützer, P., Speich, S., Karstensen, J., Labaste, M., Noisel, C., Ritschel, M., Steinhoff, T., and Wanninkhof, R.: Wintertime process study of the North Brazil Current rings reveals the region as a larger sink for CO<sub>2</sub> than expected, *Biogeosciences*, 19, 2969–2988, <https://doi.org/10.5194/bg-19-2969-2022>, 2022.
- 910 Olsen, A., Key, R. M., Van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., et al.: The Global Ocean Data Analysis Project version 2 (GLODAPv2)—an internally consistent data product for the world ocean, *Earth System Science Data*, 8, 297–323, 2016.
- Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., et al.: Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, *Nature*, 437, 681–686, 2005.
- 915 Orr, J. C., Epitalon, J.-M., Dickson, A. G., and Gattuso, J.-P.: Routine uncertainty propagation for the marine carbon dioxide system, *Marine Chemistry*, 207, 84–107, <https://doi.org/https://doi.org/10.1016/j.marchem.2018.10.006>, 2018.
- Palmiéri, J., Orr, J., Dutay, J.-C., Béranger, K., Schneider, A., Beuvier, J., and Somot, S.: Simulated anthropogenic CO<sub>2</sub> storage and acidification of the Mediterranean Sea, *Biogeosciences*, 12, 781–802, 2015.
- Park, P. K.: OCEANIC CO<sub>2</sub> SYSTEM: AN EVALUATION OF TEN METHODS OF INVESTIGATION<sup>1</sup>, *Limnology and Oceanography*,  
920 14, 179–186, <https://doi.org/https://doi.org/10.4319/lo.1969.14.2.0179>, 1969.
- Pérez, F. F., Olafsson, J., Ólafsdóttir, S. R., Fontela, M., and Takahashi, T.: Contrasting drivers and trends of ocean acidification in the subarctic Atlantic, *Scientific Reports*, 11, 1–16, 2021.
- Pfeil, B., Olsen, A., Bakker, D. C., Hankin, S., Koyuk, H., Kozyr, A., Malczyk, J., Manke, A., Metzl, N., Sabine, C. L., et al.: A uniform, quality-controlled Surface Ocean CO<sub>2</sub> Atlas (SOCAT), *Earth System Science Data*, 5, 125–143, 2013.
- 925 Rödenbeck, C., Keeling, R. F., Bakker, D. C., Metzl, N., Olsen, A., Sabine, C., and Heimann, M.: Global surface-ocean pCO<sub>2</sub> and sea–air CO<sub>2</sub> flux variability from an observation-driven ocean mixed-layer scheme, *Ocean Science*, 9, 193–216, 2013.
- Rödenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., Landschützer, P., Metzl, N., Nakaoka, S., Olsen, A., Park, G.-H., Peylin, P., Rodgers, K. B., Sasse, T. P., Schuster, U., Shutler, J. D., Valsala, V., Wanninkhof, R., and Zeng, J.: Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean pCO<sub>2</sub> Mapping intercomparison (SOCOM), *Biogeosciences*, 12,  
930 7251–7278, <https://doi.org/10.5194/bg-12-7251-2015>, 2015.
- Rose, A.: *Vision: human and electronic*, Springer Science & Business Media, 2013.
- Sabine, C. L., Key, R. M., Feely, R. A., and Greeley, D.: Inorganic carbon in the Indian Ocean: Distribution and dissolution processes, *Global Biogeochemical Cycles*, 16, 15–1, 2002.
- Sabine, C. L., Hankin, S., Koyuk, H., Bakker, D. C. E., Pfeil, B., Olsen, A., Metzl, N., Kozyr, A., Fassbender, A., Manke, A., Malczyk, J.,  
935 Akl, J., Alin, S. R., Bellerby, R. G. J., Borges, A., Boutin, J., Brown, P. J., Cai, W.-J., Chavez, F. P., Chen, A., Cosca, C., Feely, R. A., González-Dávila, M., Goyet, C., Hardman-Mountford, N., Heinze, C., Hoppema, M., Hunt, C. W., Hydes, D., Ishii, M., Johannessen, T., Key, R. M., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lourantou, A., Merlivat, L., Midorikawa, T., Mintrop, L., Miyazaki, C., Murata, A., Nakadate, A., Nakano, Y., Nakaoka, S., Nojiri, Y., Omar, A. M., Padin, X. A., Park, G.-H., Paterson, K., Perez, F. F., Pierrot, D., Poisson, A., Ríos, A. F., Salisbury, J., Santana-Casiano, J. M., Sarma, V. V. S. S., Schlitzer, R., Schneider, B.,  
940 Schuster, U., Sieger, R., Skjelvan, I., Steinhoff, T., Suzuki, T., Takahashi, T., Tedesco, K., Telszewski, M., Thomas, H., Tilbrook, B., Vandemark, D., Veness, T., Watson, A. J., Weiss, R., Wong, C. S., and Yoshikawa-Inoue, H.: Surface Ocean CO<sub>2</sub> Atlas (SOCAT) gridded data products, *Earth System Science Data*, 5, 145–153, <https://doi.org/10.5194/essd-5-145-2013>, 2013.





- 945 Sarma, V. V. S. S., Lenton, A., Law, R. M., Metzl, N., Patra, P. K., Doney, S., Lima, I. D., Dlugokencky, E., Ramonet, M., and Valsala, V.:  
Sea–air CO<sub>2</sub> fluxes in the Indian Ocean between 1990 and 2009, *Biogeosciences*, 10, 7035–7052, [https://doi.org/10.5194/bg-10-7035-](https://doi.org/10.5194/bg-10-7035-2013)  
2013, 2013.
- Skjelvan, I., Lauvset, S. K., Johannessen, T., Gundersen, K., and Skagseth, Ø.: Decadal trends in ocean acidification from the Ocean Weather  
Station M in the Norwegian Sea, *Journal of Marine Systems*, 234, 103 775, <https://doi.org/10.1016/j.jmarsys.2022.103775>, 2022.
- Steinberg, D. K., Carlson, C. A., Bates, N. R., Johnson, R. J., Michaels, A. F., and Knap, A. H.: Overview of the US JGOFS Bermuda  
Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry, *Deep Sea Research Part II: Topical Studies*  
950 *in Oceanography*, 48, 1405–1447, 2001.
- Sutton, A. J., Feely, R. A., Maenner-Jones, S., Musielwicz, S., Osborne, J., Dietrich, C., Monacci, N., Cross, J., Bott, R., Kozyr, A., et al.:  
Autonomous seawater pCO<sub>2</sub> and pH time series from 40 surface buoys and the emergence of anthropogenic trends, *Earth System Science*  
*Data*, 11, 421–439, 2019.
- Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C.,  
955 Watson, A., Bakker, D. C., 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., Delille, B., Bates, N., and  
de Baar, H. J.: Climatological mean and decadal change in surface ocean pCO<sub>2</sub>, and net sea–air CO<sub>2</sub> flux over the global oceans, *Deep*  
*Sea Research Part II: Topical Studies in Oceanography*, 56, 554–577, <https://doi.org/https://doi.org/10.1016/j.dsr2.2008.12.009>, 2009.
- Takahashi, T., Sutherland, S. C., Chipman, D. W., Goddard, J. G., Ho, C., Newberger, T., Sweeney, C., and Munro, D.: Climatological  
960 distributions of pH, pCO<sub>2</sub>, total CO<sub>2</sub>, alkalinity, and CaCO<sub>3</sub> saturation in the global surface ocean, and temporal changes at selected  
locations, *Marine Chemistry*, 164, 95–125, 2014.
- Thomsen, J., Haynert, K., Wegner, K. M., and Melzner, F.: Impact of seawater carbonate chemistry on the calcification of marine bivalves, *Biogeosciences*, 12, 4209–4220, 2015.
- Torres, O., Kwiatkowski, L., Sutton, A. J., Dorey, N., and Orr, J. C.: Characterizing mean and extreme diurnal variability of ocean CO<sub>2</sub>  
965 system variables across marine environments, *Geophysical Research Letters*, 48, e2020GL090 228, 2021.
- Uppstrom, L.: The boron/chlorinity ratio of deep-sea water from the Pacific Ocean, *Deep Sea Res.*, 21, 161–162, 1974.
- Van Heuven, S., Pierrot, D., Rae, J., Lewis, E., and Wallace, D.: CO<sub>2</sub>SYS v 1.1, MATLAB program developed for CO<sub>2</sub> system calcu-  
lations, ORNL/CDIAC-105b. Oak Ridge, TN: Oak Ridge National Laboratory, [https://cdiac.ess-dive.lbl.gov/ftp/co2sys/CO2SYS\\_calc\\_](https://cdiac.ess-dive.lbl.gov/ftp/co2sys/CO2SYS_calc_MATLAB_v1.1/)  
*MATLAB\_v1.1/*, 2011.
- 970 Woosley, R. J.: Evaluation of the temperature dependence of dissociation constants for the marine carbon system using pH and certified  
reference materials, *Marine Chemistry*, 229, 103 914, 2021.
- Zeebe, R. E. and Wolf-Gladrow, D.: CO<sub>2</sub> in seawater: equilibrium, kinetics, isotopes, 65, Gulf Professional Publishing, 2001.