# Improving coastal ocean pH estimates through assimilation of glider observations and hybrid statistical methods

Jann Paul Mattern[1], Yuichiro Takeshita[2], Carlos Rocha[1,§], and Christopher A. Edwards[1]

[1]Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, California, USA
[2]Monterey Bay Aquarium Research Institute, Moss Landing, California, USA
[§]now at Climate & Atmospheric Science, NSW Department of Climate Change, Energy, the Environment and Water, Sydney, Australia

**Correspondence:** Jann Paul Mattern (jmattern@ucsc.edu)

**Abstract.** Ocean acidification monitoring and carbon accounting require accurate estimates of marine carbonate system variables, particularly in dynamic coastal regions where observations remain sparse. This study presents an approach to improving carbonate system state estimates in the California Current System through the assimilation of underwater glider observations with both dynamical and statistical models. We implement a 4D-Var data assimilation system that jointly assimilates physical variables, chlorophyll, and glider-based pH and alkalinity data into a regional coupled physical-biogeochemical model. Our results demonstrate that joint assimilation of carbonate system variables successfully improves pH and alkalinity estimates while maintaining the quality of physical and chlorophyll estimates. Cross-validation experiments reveal that pH data assimilation typically improves estimates near the observation network, although downstream advection of increments can occasionally degrade results. We also show that hybrid estimates that combine the output of the dynamical, physical ocean model with a statistical model produce accurate carbonate system estimates without requiring a biogeochemical model. This finding suggests that physical ocean models and data assimilation systems can obtain reasonable carbonate system estimates by combining statistical methods with model estimates of temperature and salinity.

## 1 Introduction

The ocean plays a central role in modulating Earth's carbon cycle, acting as a major carbon reservoir and sink for anthropogenic emissions. Recent estimates indicate that the ocean has removed approximately one quarter of total anthropogenic carbon emissions (Friedlingstein et al., 2023). This increased uptake of atmospheric $CO_2$ results in decreased ocean pH and associated changes in carbonate system equilibrium through ocean acidification. While the coastal ocean globally acts as a sink for atmospheric $CO_2$, there is substantial regional variability (Bauer et al., 2013; Roobaert et al., 2019), with uptake largely modulated by biological carbon fixation and the biological response to climate-induced changes (Mathis et al., 2024). To understand and predict changes in ocean carbonate chemistry, the scientific community increasingly relies on coupled physical-biogeochemical ocean models (Fennel et al., 2019; Kwiatkowski et al., 2020; Fennel et al., 2022). These models combine estimates of ocean circulation dynamics with biogeochemical processes to provide state estimates of carbonate system variables and their temporal evolution.

Eastern boundary upwelling systems, such as the California Current System (CCS), represent biologically highly productive
25  regions characterized by complex coastal processes operating across multiple spatial and temporal scales (Checkley and Barth,
2009). The nearshore environment in the CCS naturally experiences low pH and reduced oxygen conditions in response to
coastal upwelling (Cheresh and Fiechter, 2020). While the southern CCS shows increased acidification consistent with global
surface ocean trends (Wolfe et al., 2023), there are substantial regional differences in alongshore mean pH and dissolved
oxygen distributions (Cheresh and Fiechter, 2020). Understanding these regional patterns and their underlying drivers requires
30  both observational data and sophisticated modeling approaches. Regional ocean models have proven successful at simulating
the physical and biogeochemical dynamics of the CCS (Veneziani et al., 2009; Fiechter et al., 2018; Deutsch et al., 2021).

Despite the ocean's crucial role in carbon cycling, marine carbonate system variables remain poorly observed. Although
millions of surface ocean $pCO_2$ measurements exist (Bakker et al., 2022), subsurface measurements of pH, alkalinity, or dissolved inorganic carbon (DIC) are comparatively sparse (Land et al., 2023). This observational gap is particularly challenging
35  in dynamic coastal regions such as the CCS, where high spatial and temporal variability demands dense sampling. Underwater
gliders have been shown to be a particularly effective autonomous platform to provide sustained, high-resolution data in a
cost-effective manner in the coastal ocean (Rudnick, 2016). For example, the California Underwater Glider Network (CUGN)
has near-continuously occupied three across-shore transects in the Southern to Central California coast since 2007, and two
more transects were added recently to the network (Rudnick et al., 2017). The CUGN gliders measure temperature, salinity,
40  water velocity, chlorophyll fluorescence, and more recently, oxygen starting in 2017 (Ren et al., 2023). There have been several
examples where underwater gliders were equipped with pH sensors (Saba et al., 2019; Takeshita et al., 2021; Hemming et al.,
2017), providing an opportunity to collect high spatiotemporal, subsurface marine carbonate system variables in the coastal
ocean. In particular for the CCS, Takeshita et al. (2021) integrated a Deep-Sea-DuraFET pH sensor (Johnson et al., 2016) onto
a Spray glider, and operated it along Line 67 in 2019, one of the transects occupied by the CUGN. Ongoing efforts to integrate
45  pH sensors into the CUGN glider fleet promises new opportunities for monitoring ocean acidification and improving model
estimates.

Data assimilation (DA) describes the process of merging models and observations, to obtain accurate and gapless estimates
of the ocean state (Edwards et al., 2015; Martin et al., 2024). The challenge of obtaining reliable marine carbonate system
state estimates has motivated various DA approaches. Several studies have explored the assimilation of carbonate system
50  variables to improve model estimates of climate-relevant parameters such as air-sea $CO_2$ exchange. Visinelli et al. (2016),
using a global state estimation setup based on the 3D-Var DA technique, found that assimilating physical observations alone
can either improve or deteriorate estimates of alkalinity and DIC depending on the region, while joint assimilation of alkalinity
and DIC improved overall estimates of air-sea $CO_2$ flux, though not consistently across all regions. Carroll et al. (2020)
employed a global coupled physical-biogeochemical model with adjoint-based physical DA and a Green's function approach
55  for biogeochemical DA to estimate carbonate system dynamics and $CO_2$ air-sea flux across seasonal to multidecadal timescales.
Turner et al. (2023) utilized ensemble optimal interpolation DA to estimate upper ocean carbon content based on temperature
and salinity observations, while Verdy and Mazloff (2017) produced biogeochemical state estimates for the Southern Ocean by
assimilating both physical data and various carbonate system and oxygen observations.

Recent work has also highlighted that carbonate system variables can be accurately estimated, particularly in the subsurface,
60   using statistical models (also known as empirical algorithms) with inputs of more commonly measured properties such as temperature, salinity, oxygen, and location (Alin et al., 2012; Bittig et al., 2018; Carter et al., 2021). These models, such as the Empirical Seawater Property Estimation Routines (ESPER; Carter et al., 2021), are trained on high-quality shipboard data (Olsen et al., 2020), using statistical regression or machine learning approaches. Using the output of dynamical ocean models as inputs to these statistical models provides an opportunity to produce gapless state estimates of the marine carbonate system.
65   These hybrid estimates, which feed the state estimates of dynamical models into statistical models, stand to benefit from DA, which aims to improve dynamical model estimates.

In this study, we integrate underwater glider-based carbonate system variables and oxygen data into a DA system based on a regional physical-biogeochemical ocean model for the CCS. In particular, we utilize ESPER-based estimates for pH, DIC, and alkalinity from 2019 using CUGN glider temperature, salinity, and oxygen as inputs. Additionally, we also utilize direct sensor
70   measurements of pH and oxygen from one transect along Line 67 (Takeshita et al., 2021). Our reference DA experiment jointly assimilates temperature, salinity, and sea level anomaly (SLA) data with satellite chlorophyll-*a* (hereafter simply chlorophyll) in a 4D-Var-based state estimation setup similar to that in Mattern et al. (2017) and Mattern et al. (2018). Through a series of experiments, we examine the effects of assimilating glider-based data alongside the reference data.

Building on previous DA work and recent glider-based pH and oxygen data, this study pursues three main goals: First,
75   we examine the impact of jointly assimilating carbonate system variables (primarily pH and alkalinity) and oxygen with physical and chlorophyll observations. Second, we assess the influence of assimilating different data subsets on the model's state estimates, examining the relative importance of various observation types and their interactions in the assimilation system. Third, we investigate whether integrating ROMS model temperature and salinity estimates with the ESPER statistical model can produce useful hybrid estimates of carbonate system variables, potentially offering a practical approach for carbonate
80   system state estimation based on physical ocean models and DA systems. With our DA framework, we ultimately aim to address the need for improved coastal carbonate system monitoring and the broader challenge of developing efficient methods for ocean acidification tracking in dynamic coastal environments.

## 2   Methods

### 2.1   Model domain and physical setup

85   Our model is based on version 3.7 of the Regional Ocean Modeling System (ROMS; Haidvogel et al. (2008)), a primitive equation, hydrostatic, free surface regional ocean physical circulation model. The model domain is designed to simulate physical and biogeochemical dynamics in the CCS. It extends from 30°N to 48°N, and longitudinally from the U.S. west coast to 134°W, with a horizontal resolution of 0.1° × 0.1° (resulting in a ≈10 km resolution) The model is vertically divided into 42 terrain-following layers with a higher resolution near the ocean surface. Our implementation uses the Coupled Ocean Atmo-
90   sphere Mesoscale Prediction System (COAMPS; Doyle et al., 2009) for physical forcing at the surface (wind, longwave solar radiation, air temperature, pressure, and humidity). Boundary conditions for the physical model are provided by Copernicus

Global 1/12° reanalysis (GLORYS; Jean-Michel et al., 2021). The GLORYS product is further used to nudge model temperature and salinity toward their reanalysis values in a 1° wide "sponge" region along the open model boundaries. This physical model setup closely follows that in previous studies, such as Veneziani et al. (2009); Mattern et al. (2017).

## 2.2 The NEMUCSC biogeochemical model

As a biogeochemical model, we use the NEMUCSC model, an extension of the North Pacific Ecosystem Model for Understanding Regional Oceanography (NEMURO; Kishi et al., 2007). Our NEMUCSC configuration has been extended from the setup presented in Mattern et al. (2017) with carbon and oxygen cycling and phytoplankton photoadaptation (i.e., dynamic nitrogen to chlorophyll ratios for phytoplankton, which we detail below).

The base NEMUCSC model accounts for a range of biogeochemical constituents and interactions in the ecosystem, including: (1) three limiting macronutrients (nitrate, ammonium, and silicate); (2) two phytoplankton functional types (nanophytoplankton and diatoms); (3) three zooplankton size classes (microzooplankton, mesozooplankton, and predatory zooplankton); and (4) three detritus pools (dissolved and particulate organic nitrogen and particulate silica). The NEMUCSC carbon extension, presented in Cheresh and Fiechter (2020) and further refined in Green et al. (in prep.), adds dissolved inorganic carbon (DIC), total alkalinity, and calcium carbonate variables. It computes ocean pH and pCO$_2$ based on the OCMIP carbonate chemistry. The carbon model includes the major fluxes influencing DIC and alkalinity, with DIC decreased by primary production and calcium carbonate formation and increased by respiration, remineralization, and calcium carbonate dissolution. Alkalinity is modulated by processes such as nitrification, calcium carbonate formation, new primary production, and calcium carbonate dissolution (Fig. 1).

The second extension to the NEMUCSC model adds a chlorophyll variable for each of the two phytoplankton variables. Nitrogen is the base currency in NEMUCSC, so that diatoms and nanophytoplankton are represented as nitrogen-based variables. The photoadaptation model presented in Jackson et al. (2017) is used to dynamically adjust the stoichiometric ratio of nitrogen to chlorophyll for the two phytoplankton functional types represented in the model, assuming a constant nitrogen-to-carbon ratio based on the Redfield ratio.

We further expanded the code required by the DA (NEMURO's tangent-linear and adjoint code; see Mattern et al. (2017)) to fully support the extended model and permit the assimilation of chlorophyll, oxygen and carbonate system variables. For DA, we added a "total chlorophyll" variable which is the sum of diatom chlorophyll and nanophytoplankton chlorophyll, so that chlorophyll observations can be directly assimilated, informing both phytoplankton functional types. Following Song et al. (2012); Mattern et al. (2017), chlorophyll is log-transformed in the DA procedure. A log-transformation is not applied to pH, alkalinity, DIC, oxygen or the physical variables for which we also assimilate data in this study.

For the carbonate system variables DIC and alkalinity, as well as oxygen, nitrate and silicate, boundary conditions are created using statistical relationships from ESPER, based on temperature and salinity boundary values. Following a comparison with climatological nutrient values from the World Ocean Atlas (Garcia et al., 2024), nitrate boundary values in the mixed layer were reduced to match World Ocean Atlas values, and to prevent unrealistically high nitrate input through the boundaries. To further prevent unrealistic values for other variables, boundary conditions for the remaining BGC variables are set to low
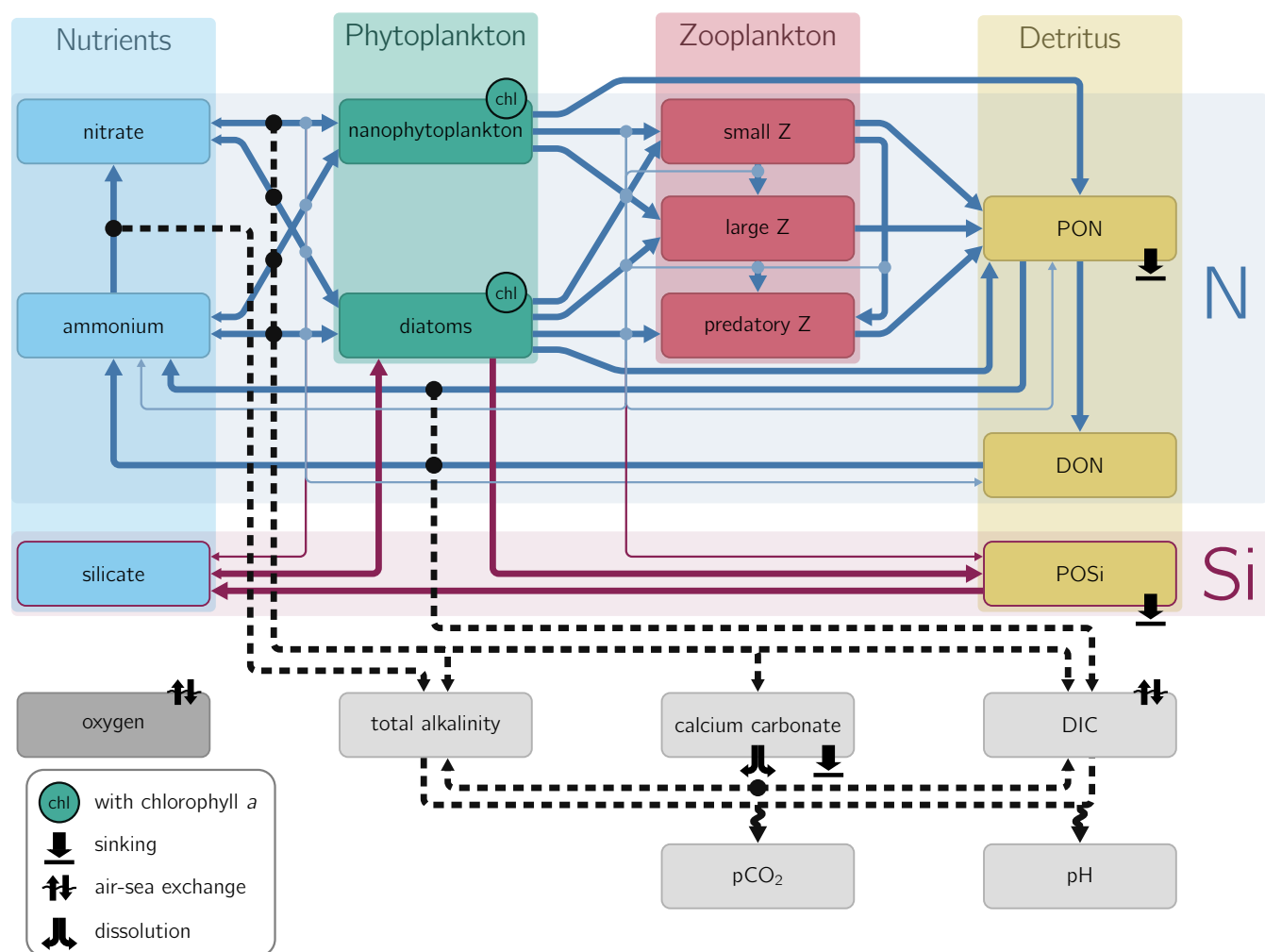
**Figure 1. NEMUCSC model diagram:** Outline of the NEMUCSC model variables and pathways connecting them. Nitrogen and silicate-based variables are outlined in blue and red, respectively, carbonate system variables are shown in light gray. Pathways between nitrogen and silicate-based variables, indicated by thick solid arrows for major processes (e.g. grazing) and thinner solid arrows for minor processes (e.g. particulate matter exudation due to sloppy/inefficient grazing) transfer matter between variables. In contrast, matter is not transferred between carbonate system variables; instead, biogeochemical processes affecting a carbonate system variable lead to an increase or decrease of that variable, the connection between processes and affected carbonate system variable is shown by black dashed arrows. Pathways affecting model oxygen are not shown.

values that roughly match model estimates near the boundaries. For the year of our experiments, 2019, atmospheric pCO$_2$ is set to a constant value of 400 ppm.
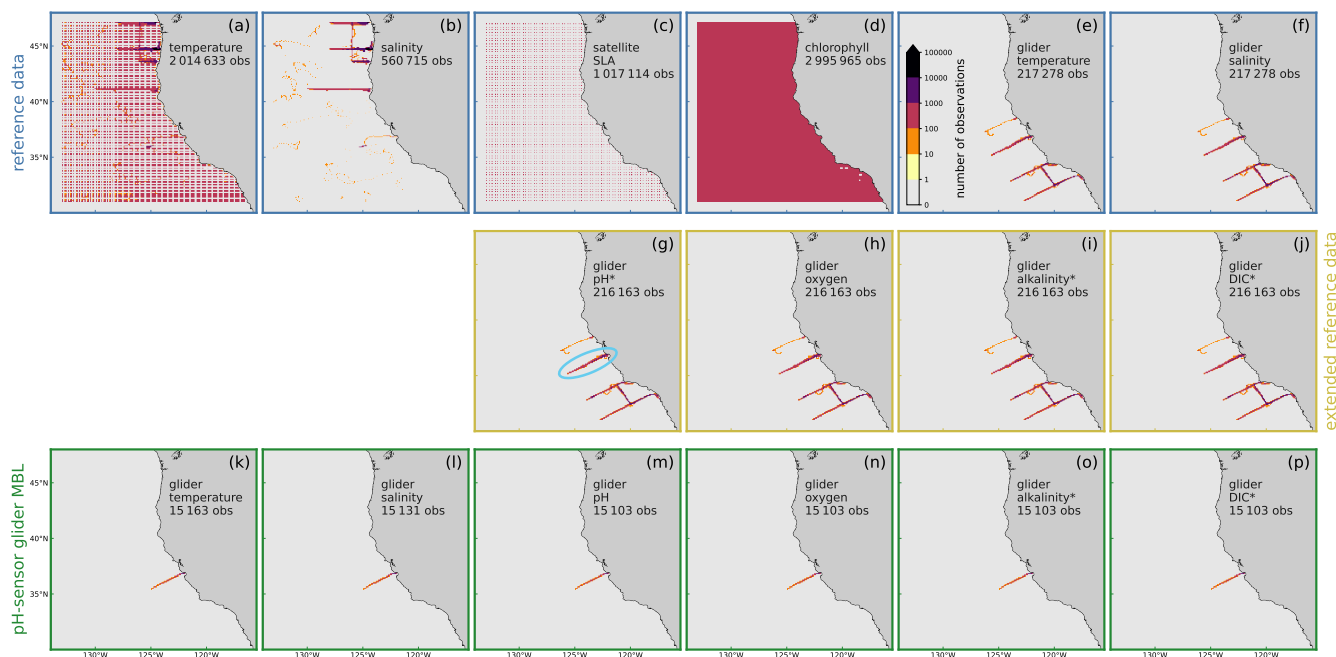
**Figure 2. Histogram of observation locations:** 2-dimensional histogram showing the number of observations and their horizontal location on the model grid for each observation type. The observations are divided into 3 categories: (a-f) The reference data contains mostly remotely sensed observations, augmented by glider- and float-based observations of temperature and salinity. (g-j) The extended reference further includes glider-based data for pH, oxygen, alkalinity and DIC. (k-p) The pH-sensor glider dataset contains glider data from independent deployments. Asterisks (*), denote data that was not directly measured, but obtained from other measurements using statistical relationships. The light blue ellipse in (g) marks the Line 67.

## 2.3 Data used in the assimilation and verification

In this study, we use a variety of satellite, glider, and float observations, complemented by estimated data derived from statistical relationships using the ESPER-LIR algorithm. We conducted a series of experiments in which various data subsets were excluded from the assimilation process to examine their influence on model state estimates. Furthermore, we used model state estimates with and without DA and ESPER-LIR to derive hybrid (dynamical model + statistical model) state estimates.

The datasets are classified into three broad categories:

1. **Reference dataset**: this represents what is typically assimilated in a coupled physical-BGC model, encompassing:

   – Temperature data from satellites

   – In situ observations of temperature and salinity from gliders, including gliders from the California Underwater Glider Network (CUGN), and floats

   – Satellite-based sea level anomaly (SLA) data

**Table 1.** The data used for assimilation and verification.

| observed variables | data source | DOI/URL | category |
| --- | --- | --- | --- |
| temperature (T) | Ifremer SST multi-sensor L3 | https://doi.org/10.48670/moi-00164 | reference |
| T + salinity (S) | Argovis subset of ARGO data | https://argovis.colorado.edu | reference |
| T + S | California Underwater Glider Network | https://spraydata.ucsd.edu | reference |
| T + S | IOOS glider data | https://gliders.ioos.us | reference |
| chlorophyll-*a* | CCI chlorophyll-*a* v6 | https://www.oceancolour.org | reference |
| SSH | Copernicus Gridded L4 SSH | https://doi.org/10.48670/moi-00148 | reference |
| oxygen | California Underwater Glider Network | https://spraydata.ucsd.edu | ext. ref. |

    – Chlorophyll data derived from satellite ocean color

2. **Extended reference dataset**: this supplements the reference data by incorporating

    – Oxygen observations from the CUGN glider deployments

    – Estimated values of pH, alkalinity, and DIC obtained from ESPER-based estimates with inputs of glider temperature, salinity, and oxygen observations

3. **pH-sensor glider line dataset**: this dataset is based on several deployments of a glider with an in situ pH sensor for Line 67, extending westward out of Monterey Bay. It includes:

    – Temperature, salinity, pH, and oxygen observations

    – Estimated values for alkalinity using ESPER and estimated DIC by combining measured pH and estimated alkalinity

All data used in our DA experiments are freely available. The sources for the datasets and their association with the categories are listed in Table 1.

In our DA experiments, we assimilate so-called super-observations, which are observations that are averaged based on the model grid and time stepping so that there is at most one observation of the same type in each model grid cell and time step. The use of super-observations is standard in DA applications (Martin et al., 2024); for dense data sets, such as glider-based observations, the process of creating super-observations can reduce the number of data points considerably. However, in quantitative evaluations that use glider-based observations, we use the original observations, not the super-observations. We adopt this approach for two reasons: first, we are more interested in improving the fit to the full set of observations rather than the model grid-dependent aggregated data. Second, for visualization purposes, a denser set of observations better reveals observed features. Interestingly, the overall model fit to the original glider observations is typically better than for the super-observations, a phenomenon we discuss in more detail in Section 4. Because only super-observations are assimilated, we often

**Table 2.** Experiments and assimilated data

|  | ref data | CUGN pH$^*$, alk$^*$ | pH-sensor glider line T, S, pH, alk$^*$ | CUGN oxygen |
|---|---|---|---|---|
| free simulation |  |  |  |  |
| ref DA | ✓ |  |  |  |
| DA exp 1 | ✓ | ✓ |  |  |
| DA exp 2 | ✓ | ✓ | ✓ |  |
| DA exp 3 | ✓ | ✓ |  | ✓ |
| DA exp 1.1 | ✓[1] | ✓[1] |  |  |
| DA exp 1.2 | ✓[2] | ✓[2] |  |  |
| DA exp 1.3 | ✓[3] | ✓[3] |  |  |

$^*$Data not directly measured, estimated based on statistical relationships. [1]Excluding CUGN L67Mar. [2]Excluding CUGN L67Jul.
[3]Excluding CUGN L67Oct.

160 drop the "super" prefix and refer to the assimilated data simply as "observations" or "data". When needed, we use the term
super-observations to avoid ambiguity.

## 2.4 Data assimilation experiments

In this study, we conducted several DA experiments to systematically investigate the impact of assimilating different observation subsets. Table 2 provides an overview of all experiments and the data assimilated in each.

165 To establish a baseline, we use two reference simulations: a non-assimilative "free simulation" and a reference DA experiment. The free simulation does not assimilate any observations and is solely driven by the coupled physical-biogeochemical model. Our reference DA experiment represents a typical joint physical-biogeochemical DA approach. It assimilates remotely sensed and in situ temperature and salinity data, satellite sea level anomaly (SLA), and satellite-derived chlorophyll. This experiment serves as a benchmark for current physical-biogeochemical DA without any carbonate system observations.

170 Building upon the reference DA, we conducted three main experiments that progressively incorporate additional observation types: DA exp 1 assimilates all reference data plus CUGN pH and alkalinity estimates. This experiment examines the impact of jointly assimilating carbonate system variables with the standard physical and chlorophyll data.

DA exp 2 builds on DA exp 1 by additionally assimilating pH-sensor glider line temperature, salinity, pH, and alkalinity data. This experiment aim is to examine the model estimates achievable through direct assimilation of the pH-sensor glider
175 data , compared to the assimilation of neighboring CUGN pH and alkalinity in DA exp 1.

DA exp 3 assimilates all reference data plus CUGN pH, alkalinity, and oxygen data. This experiment investigates the impact of including oxygen observations in the assimilation framework.

To assess the spatial influence of the pH DA and validate our approach, we conducted three cross-validation experiments: we focused on three CUGN glider transects out of Monterey Bay along Line 67 (see highlighted line in Fig. 2g) in different
180 seasons (starting in late March, July, and October 2019). During each transect, the glider moved westward out of Monterey

Bay, covered the full extent of the line, and returned nearshore. We denote these three lines as CUGN L67Mar, CUGN L67Jul, and CUGN L67Oct (started in March, July, and October 2019, respectively) and performed an experiment excluding the data for each line. DA exp 1.1, 1.2, and 1.3 assimilate all data from DA exp 1 except CUGN L67Mar, L67Jul, or L67Oct data, respectively. These cross-validation experiments allow us to evaluate the model's ability to estimate pH and other variables at 185 locations where observations are withheld from the assimilation.

In addition to these direct assimilation experiments, we create hybrid estimates by using dynamical ocean model outputs as inputs to the ESPER statistical model. ESPER can provide estimates of various biogeochemical seawater properties, including pH, alkalinity, DIC, and oxygen. It generates these estimates based on location information, temperature and salinity, and optionally, additional inputs such as oxygen. We can use the model estimates of temperature and salinity as input to ESPER to 190 obtain pH, alkalinity, DIC, and oxygen estimates. These hybrid estimates, combining the dynamical model with the statistical model, may benefit from physical DA as it improves the temperature and salinity estimates that enter ESPER. We created hybrid estimates based on

1. the free simulation (ESPER based on temperature and salinity only),

2. the reference DA (ESPER based on temperature and salinity only),

195    3. DA exp 3 (ESPER based on temperature, salinity, and oxygen).

This experimental design allows us to systematically evaluate the impact of different observation types on model state estimates and to compare direct assimilation with hybrid approaches for generating carbonate system estimates.

## 2.5 Observation and background error values

The 4dVar DA system requires the specification of observation error and background error values, the latter to quantify the 200 model uncertainty associated with the background model state, prior to performing DA. Previous experience working with joint 4dVar DA for physical and chlorophyll-*a* observations, and differences between the observed variables in the carbonate system submodule and the rest of the model suggest different approaches for specifying error values for these two parts of the model.

Mattern et al. (2018) outlines an iterative technique to adjust observation error and background error values to improve the 205 consistency of prior and posterior error values. The adjustments were shown to have a positive effect on 4dVar performance diagnostics and model performance metrics, such as the forecasting skill. This technique, which we hereafter refer to as a fixed point iteration, or "FPI", can only be applied to the background error values of directly observed variables or variables with a simple relationship to assimilated observations. In the physical model, the temperature, salinity, and SSH variables are directly observed, which means that the observations of temperature, salinity, and SSH correspond directly to these model variables. 210 Chlorophyll-*a* is also directly observed, but the value of the chlorophyll-*a* variable is a function of, and mostly determined by the values of the model's two phytoplankton variables, diatom and nanophytoplankton. By assuming a constant phyto-plankton nitrogen:chlorophyll-*a* ratio, we can create a simple (linear) relationship between the phytoplankton variables and the

**Table 3.** The observation error values used in the assimilation.

| observed variable | subset | observation error | observation error origin |
|---|---|---|---|
| temperature | all in situ | 0.28°C | FPI |
| temperature | satellite | 0.25-1.16°C | dataset values modified by FPI |
| salinity | all in situ | 0.15 | FPI |
| SLA | satellite | 0.06 m | FPI |
| chlorophyll-*a* | satellite | 58.3% | FPI |
| pH | extended reference stat. estimate | 0.05 | dataset values |
| pH | Line 67 in situ | 0.013 | dataset values |
| alkalinity | extended reference stat. estimate | 15.0 mmol m$^{-3}$ | dataset values |
| alkalinity | Line 67 stat. estimate | 7.0 mmol m$^{-3}$ | dataset values |
| DIC | extended reference stat. estimate | 20.0 mmol C m$^{-3}$ | dataset values |
| DIC | Line 67 stat. estimate | 8.0 mmol C m$^{-3}$ | dataset values |
| oxygen | all in situ | 3.0 mmol m$^{-3}$ | dataset values |

chlorophyll-*a* observations. This approach permits us, starting from initial estimates, to utilize the FPI to adjust temperature, salinity, SSH, and chlorophyll-*a* observation error values and the background error values for the model variables associated with these observations (see Mattern et al. (2018) for details).

Because the relationship between observations of the carbonate system and model variables is often more complex, the FPI cannot easily be applied to variables of the carbonate system. For example, model pH is a complex, non-linear function of alkalinity, DIC, temperature, and salinity. Thus, we rely on our initial estimates of observation errors and background error values for the carbonate system variables.

For observation error values, we used the uncertainty values provided with the datasets when available and constant errors otherwise. Specifically, we started with 0.1°C, 0.05, 10 cm, and 30% (relative error) as the observation error estimates for in situ temperature, in situ salinity, SSH, and chlorophyll, respectively, when no uncertainty values were provided by the datasets. These values were adjusted using the FPI described above; the resulting error values used in our experiments are listed in Table 3 alongside the error values used for the other observations that were not modified.

The prescribed background error values (specifying the diagonal entries of the background error covariance matrix) for all model variables are assumed to be constant at the surface and decrease with depth following an S-shaped logistic function (for all variables except SLA, which has no depth-dependence). Below 200 m, where oceanic variability is weaker, background error values are set to approximately 10% of their magnitude at the surface (following the approach in Mattern et al. (2018)).

The background error values for temperature, salinity, SSH, diatom, and nanophytoplankton are set using the FPI. For the carbonate system variables and oxygen, we use a long, non-assimilative model simulation to set the surface value: for each variable, we assume that all daily average surface values are contained within a window of ± 3 standard deviations from its mean. Thus, we take the difference between the maximum daily average surface value in the simulation and its minimum and

**Table 4.** The background error values used in the assimilation. All background error values are constant horizontally and highest at the surface, declining to 10% of their surface at depth below 200 m (except for SLA which is only defined at the surface). Background error values for diagnostic variables, such as pH, and semi-diagnostic variables, such as diatom chlorophyll-*a*, have little impact on the DA result and are not shown.

| model variable(s) | surface background model error value | origin of error value |
|---|---|---|
| temperature | 0.4°C | FPI |
| salinity | 0.03 | FPI |
| SLA | 0.036 m | FPI |
| nanophytoplankton and diatoms | 0.63 mmol m$^{-3}$ | FPI |
| alkalinity | 20.2 mmol m$^{-3}$ | 3 $\sigma$-approach |
| DIC | 47.8 mmol m$^{-3}$ | 3 $\sigma$-approach |
| calcium carbonate | 0.33 mmol m$^{-3}$ | 3 $\sigma$-approach |
| oxygen | 30.5 mmol m$^{-3}$ | 3 $\sigma$-approach |
| u and v velocities | 0.03 m s$^{-1}$ | low value for unobserved |
| all other | 0.3 mmol m$^{-3}$ | low value for unobserved |

divide it by 6 to obtain the standard deviation (this is referred to as "3 $\sigma$-approach" in Table 4). The background error values for the other model variables are set to low values to prevent large increments to unobserved variables. All surface background error values used in our experiments are listed in Table 4.

## 2.6 Considerations for assimilating pH data

In a state estimation setting, assimilating pH data (observed or estimated) alone can deteriorate carbon state estimates. This issue arises because, in the model, pH is primarily a function of, and calculated from model estimates of DIC and alkalinity. Different combinations of alkalinity and DIC can yield the same pH value, resulting in model alkalinity and DIC estimates being underdetermined by pH data, analogous to temperature and salinity estimates being underdetermined by density observations in a physical assimilation setting.

In our initial experiments, where we assimilated pH observations alone, state estimates were moved to points in alkalinity-DIC space with the correct pH value. However, the resulting estimates of alkalinity and DIC were largely based on their initial estimates and the prescribed background model error values, rather than reflecting the true state of the system (see also Fennel et al. (2023)). To move state estimates to the correct point in alkalinity-DIC space, we found it necessary to assimilate two out of the three properties: pH, alkalinity, and DIC.

For this study, we elected to assimilate pH, as it was directly measured on Line 67, and estimated alkalinity because its uncertainty is lower than that of DIC. This choice means that we are using and relying on statistical estimates of alkalinity in the DA procedure, even when pH is directly observed. While this approach introduces the reliance on statistically estimated

250   alkalinity, alkalinity can be estimated with an accuracy of $< 7 \; \mu$mol/kg across the CCS, even near the coast (Takeshita et al., 2021).

## 3   Results

This study is guided by three primary goals: (1) Examine the impact of jointly assimilating carbonate system variables (primarily pH and alkalinity) and oxygen with physical and chlorophyll observations. (2) Assess the influence of various data
255   subsets on the DA state estimates. (3) Determine whether the application of the dynamical ROMS model state estimates with the statistical ESPER estimates based on ROMS temperature and salinity can improve the precision of carbonate system state estimates, providing carbonate system state estimates without a carbonate system model or carbonate system DA. These goals inform the design and interpretation of the experiments presented below.

### 3.1   Free and assimilative reference simulations

260   Our DA experiments differ only in the subsets of data being assimilated (see Section 2.4). To establish a baseline, we use two reference simulations: a reference DA experiment and a free simulation that assimilates no data.

We first examine the root mean squared error (RMSE) between model and data for each dataset introduced in Section 2.3. Comparing the two experiments reveals that, similar to previous studies jointly assimilating chlorophyll and physical data (e.g. Mattern et al., 2017), the DA substantially improves the fit to the assimilated data compared to the free simulation (compare
265   gray and blue bars in Fig. 3). However, this reduction in error for physical and chlorophyll data (first row in Fig. 3) does not translate to improved fit for glider observations of carbonate system variables (pH, alkalinity, DIC) or oxygen (second row in Fig. 3). Notably, there is no improvement in the reference DA experiment even though it assimilates temperature and salinity data from the same CUGN glider lines associated with the carbonate system variables.

A closer examination of a representative glider line reveals the spatial structure of the model-observation differences for
270   pH. We focus on a CUGN Line 67 glider transect out of Monterey Bay (see highlighted line in Fig. 2g) starting in late March 2019, which we refer to as CUGN L67Mar (Fig. 4). The discrepancies between pH observations and free simulation model estimates along this line are highest between 50 m and 200 m depth (Fig. 4d, e, f), corresponding to density surfaces of 1025.5-1027 kg m$^{-3}$. Near the surface and at depths below 300 m, the free simulation shows a positive bias, overestimating pH by $\approx 0.1$. Assimilation of the reference data, including temperature and salinity data from this line, does not remove the surface
275   or deep pH bias, and results in minor shifts in the largest model-observations differences at intermediate depths between 50 m and 200 m (compare Fig. 4f and h). The improvement in pH estimates, measured by the reduction in the absolute model-observation difference of the reference DA in comparison to the free simulation (Fig. 4i) confirms that the largest change in model pH through the assimilation of the reference data occurs in the same depth range (50 m to 200 m), where the reference DA leads to positive and negative changes in pH estimates.
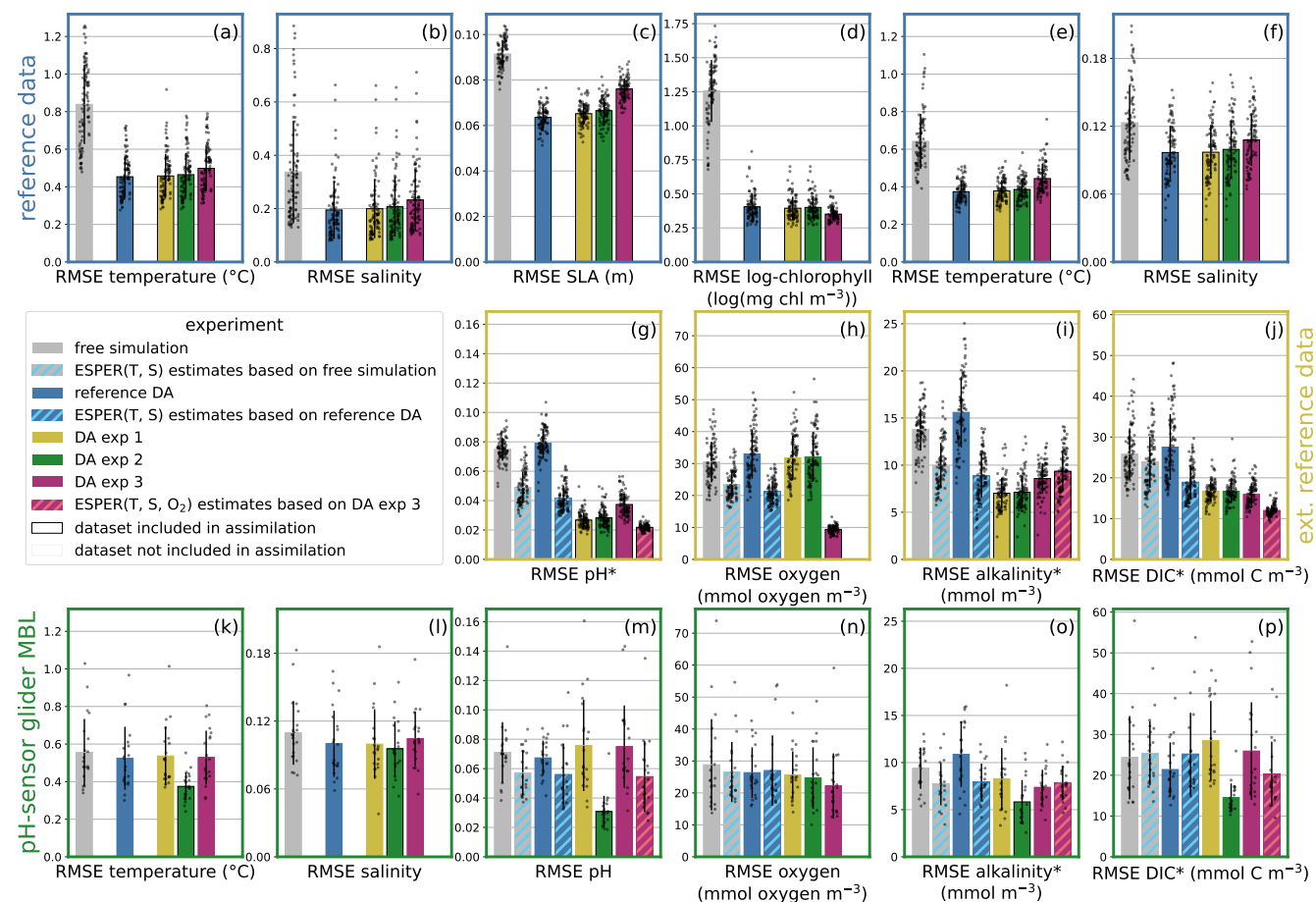
**Figure 3. RMSE for the main experiments:** Each panel corresponds to a dataset (displayed in Fig. 2 in the same panel layout), the bars represent model experiments: small black dots display the RMSE of a simulation for that dataset, for each cycle which includes data for the dataset. The height of the bar displays the average cycle RMSE, and the black line on each bar shows the mean ± standard deviation of the cycle RMSE. Black outlines of bars indicate that the dataset was assimilated in the experiment.

280    In our state estimation setup, the joint assimilation of temperature, salinity, SLA, and chlorophyll data has negligible impact on estimates of pH, oxygen, alkalinity, or DIC. This finding aligns with previous state estimation studies reporting little impact of physical DA on chlorophyll estimates, which we will discuss further in Section 4.

## 3.2    Assimilating pH and alkalinity data

Because the joint physical and chlorophyll DA showed no consistent improvements in pH estimates, we next examine the
285    simulation that assimilates CUGN-based pH and alkalinity estimates in addition to the reference data, in the experiment referred to as "DA exp 1". The joint assimilation of pH and alkalinity estimates with physical and chlorophyll data leads to significant
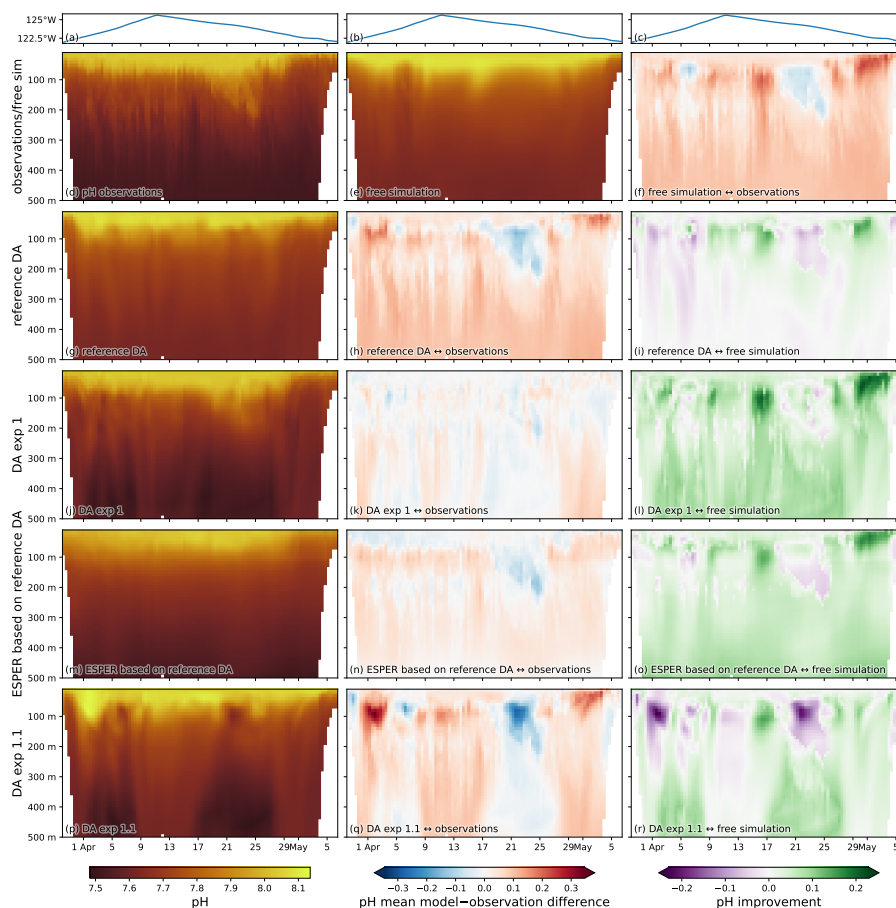
**Figure 4. Glider pH transect for CUGN L67Mar:** The glider-based pH observations and model pH estimates at the observation locations for a glider transect along Line 67 started in late March 2019. (a-c) The longitude of the glider, indicating its position along the east-west line for each date. (d) Glider pH observations binned on a date-depth grid focused on the top 500m. The model pH estimates on the same grid for the (e) free simulation, (g) reference DA, (j) DA exp 1, (m) hybrid estimate using ESPER based on the reference DA, and (p) DA exp 1.1. (f, h, k, n, q) The model-observation difference for pH along the transect for all 5 pH estimates. (i, l, o, r) The improvement of the estimates in comparison to the free simulation (the absolute value of the model-observation difference for the free simulation minus the absolute value of the model-observation difference for the DA-based or hybrid estimate). Positive (green) values indicate a reduction in model misfit of the DA-based or hybrid estimate in comparison to the free simulation.

improvement in the model carbonate chemistry parameters. DA exp 1 more than halves the average RMSE of the CUGN pH data compared to the free simulation or the reference DA, while achieving a similar fit to the reference data temperature, salinity, SLA and chlorophyll as the reference DA (Fig. 3). The improvement in pH is accompanied by improved fit to CUGN alkalinity and DIC data (Fig. 3i, j), with little change in oxygen estimates (Fig. 3h).
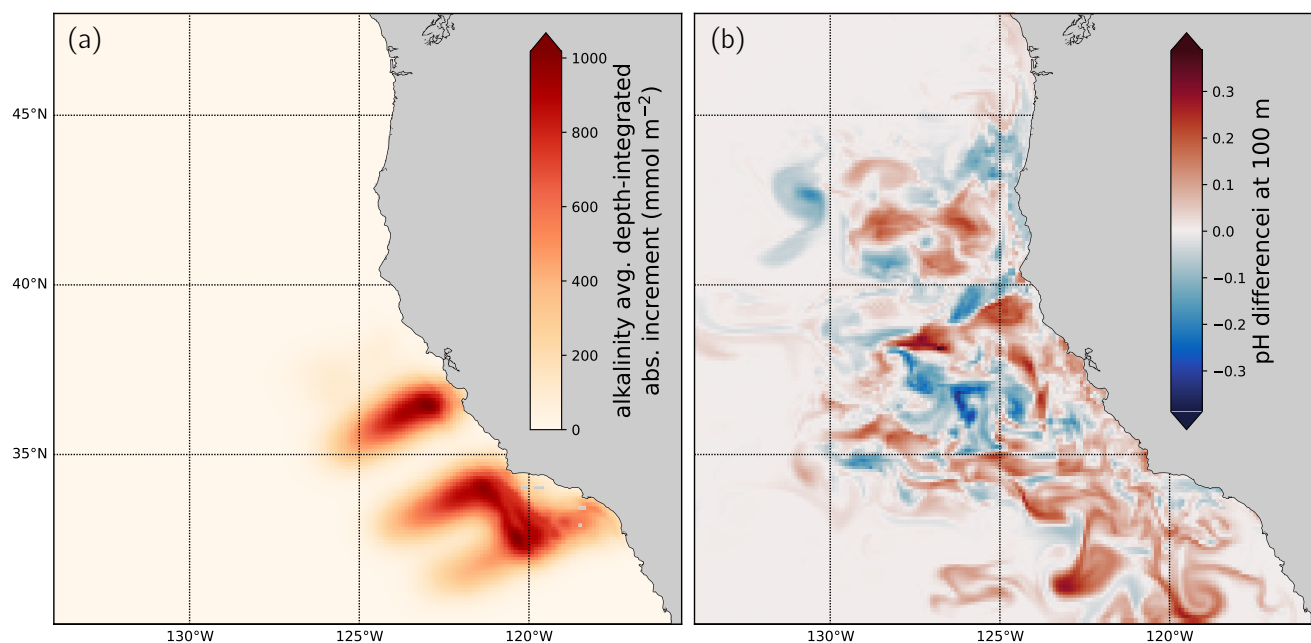
**Figure 5. Impact of the pH DA:** (a) The average depth-integrated absolute increment in alkalinity across all 91 DA cycles in DA exp 1. (b) The difference in pH between the reference DA (not assimilating pH data) and DA exp 1 at 100 m depth at the end of the 91 DA cycles on 2019-12-31.

Examining the CUGN L67Mar transect (Fig. 4j-l), the assimilation of pH data leads to overall improvement in pH estimates: surface and deep biases shown in the reference DA are removed, and larger pH discrepancies at intermediate depths disappear.

In the results above, we have only examined the fit to data that has been assimilated. To better understand the spatial influence of pH DA, we examine the increments in DA exp 1 to the carbonate system variables. When only assimilating physical and chlorophyll data, the DA system does not create corrective increments to the carbonate system variables (as these have no direct influence on the physical or chlorophyll state estimates; compare Fig. 1). That is, in the reference DA, the carbonate system variables are only affected *indirectly* by the DA through increments in the physical and nitrogen-based variables and subsequent changes in physical and biogeochemical model dynamics. In DA exp 1, however, the assimilation creates significant increments to alkalinity and DIC to improve the fit to the CUGN pH and alkalinity observations.

The average depth-integrated absolute alkalinity increment (Fig. 5a) illustrates the horizontal extent of influence of the pH assimilation, clearly outlining the 2019 CUGN glider lines that were assimilated (compare Fig. 2g, i). Each increment represents the difference between the model state before and after DA at the start of each cycle, which is influenced by the 50 km horizontal length scale prescribed in the DA setup. Thus, it does not capture the subsequent effect of DA due to the model dynamics (the cumulative effect of the increments on the physical and biogeochemical model dynamics through each cycle).

To assess the downstream effect of pH assimilation, we examine the difference in pH estimates at 100 m depth between DA exp 1 and the reference DA at the end of the experiments (2019-12-31). After a full year of assimilating data, the pH difference at 100 m depth has spread from the pH observation locations along the coast and offshore (Fig. 5b). The snapshot of pH difference reveals complex structures, with areas of positive pH change bordering water masses with negative change.

310   In Sections 3.4 and 3.5, we examine if the pH DA improves the fit to pH data that has not been assimilated.

## 3.3   Hybrid estimates based on ESPER

Hybrid estimates use the output of the dynamical model as input to the statistical ESPER model (see Section 2.4). As DA can lead to substantial improvements in dynamical model estimates, it is likely to produce better hybrid estimates.

Because the temperature and salinity model estimates do not appear to benefit from the assimilation of carbonate system

315   data (Fig. 3e, f), we only produce ESPER estimates based on the output of the free simulation and the reference DA. We do not use model oxygen to inform ESPER estimates because model initial conditions for oxygen are based on ESPER estimates and DA experiments have shown little improvement in oxygen estimates compared to the free simulation (compare Fig. 3h). Instead, here we use ESPER based on temperature and salinity only, and examine the assimilation of oxygen data and ESPER estimates based on assimilated temperature, salinity, and oxygen in Section 3.6.

320   Using model temperature and salinity, we generate ESPER-based estimates throughout the entire model domain. Here, we focus on ESPER estimates of pH, alkalinity, and DIC at the observation locations, that is, at the time and place glider observations were recorded. Comparing these values to the CUGN data reveals that hybrid estimates have a better fit than model-based estimates for pH, alkalinity, and DIC (Fig. 3g, i, j), except for DA exp 1 which directly assimilates CUGN pH and alkalinity data. For oxygen, hybrid estimates create the closest fit to CUGN observations among all estimates considered

325   so far. In all cases, ESPER estimates based on the reference DA outperform those based on the free simulation, benefiting from improved temperature and salinity estimates obtained through physical DA.

The hybrid pH estimates for ESPER based on the reference DA along CUGN L67Mar (Fig. 4m-o), show a reduced misfit to data compared to the free simulation and reference DA. The largest misfit occurs at depths around 100 m, corresponding to the $1026 \text{ kg m}^{-3}$ density surface.

330   In summary, hybrid estimates based on model temperature and salinity, combined with the ESPER statistical model, produce some of the best carbonate system estimates in our study, unless the carbonate system data itself is assimilated. Notably, these hybrid estimates do not require a biogeochemical ocean model and can be obtained from a physical ocean model and physical DA, an aspect we examine in more detail in Section 4.

## 3.4   Evaluation using independent Line 67 data

335   In previous sections, we examined model-based and hybrid estimates for data that was assimilated in some experiments and where pH and alkalinity were not directly measured but estimated using a statistical model. Here, we evaluate the same estimates against a new glider-based dataset that is not part of the CUGN data, contains direct pH measurements along with temperature, salinity, and oxygen, and was not assimilated in the experiments introduced so far. This dataset, referred to as the

pH-sensor glider line data, was recorded during several MBARI glider deployments. These deployments spatially overlap with
340   the CUGN Line 67 (see Fig. 2) and are thus in the region strongly influenced by the CUGN pH DA (see Fig. 5).

For comparison with the pH-sensor glider line data, we introduce an additional experiment, DA exp 2, which assimilates
pH-sensor glider line temperature, salinity, pH, and alkalinity data, as well as reference data and CUGN pH and alkalinity
data. DA exp 2 serves two main functions: to determine the estimates for the pH-sensor glider line achievable through direct
assimilation and to demonstrate that CUGN pH data and measured pH-sensor glider line pH data can be jointly assimilated
345   without deteriorating estimates for either dataset. To run DA exp 2, we rely on statistical estimates of alkalinity for the pH-
sensor glider line, jointly assimilated with measured pH data, and include statistical estimates of DIC in the pH-sensor glider
line data for comparison.

We first examine the 4-day RMSE values of model and hybrid estimates with respect to the pH-sensor glider line data
(Fig. 3k-p). For temperature and salinity, DA experiments perform similarly to the free simulation, with only DA exp 2 showing
350   improvement. Assimilation of large temperature and salinity datasets, including CUGN data, does little to improve overall
estimates at the pH-sensor glider line.

These results from physical data also apply to the carbonate system variables. For pH, DA exp 1 performs slightly worse on
average than the reference DA and free simulation, with only hybrid estimates and direct assimilation (DA exp 2) performing
better. Results for alkalinity and DIC are similar to pH; for oxygen, all estimates perform comparably, and DA-based estimates
355   only slightly outperform the free simulation.

For pH, the largest misfit between model estimates and pH-sensor glider line observations occurs near the surface and
at intermediate depths down to 150m, with no deep pH bias for simulations assimilating CUGN pH data (not shown). The
structure of pH differences between estimates from the reference DA and DA exp 1 (Fig. 5b) provides context for why the DA
cannot substantially improve intermediate depth estimates along glider Line 67 extending out of Monterey Bay. The region
360   is highly active in terms of physical circulation and DA impact, with large pH changes occurring on relatively small spatial
scales. We conclude that our DA requires spatially or temporally denser observations to inform the model about the evolving
pH field, and only the direct assimilation of pH-sensor glider line pH observations significantly reduces pH misfit in the top
150m.

## 3.5   Cross-validation experiments

365   To better understand model pH estimates for non-assimilated data, we performed three additional cross-validation experiments.
In these experiments, different subsets of CUGN pH and alkalinity data were withheld from assimilation. We focus on the Line
67, and three glider transects in different seasons: CUGN L67Mar, CUGN L67Jul, and CUGN L67Oct (CUGN L67Mar is
used in Fig. 4). For each line, we created an experiment that assimilates all data from DA exp 1 (reference data and CUGN
pH, alkalinity), except for the CUGN temperature, salinity, pH, and alkalinity data associated with that specific transect. These
370   three new experiments are denoted DA exp 1.1, 1.2, and 1.3 (see Table 2).

We examine the RMSE of DA exp 1.1, 1.2, 1.3 in comparison to DA exp 1, the reference DA, and free simulation for
the three datasets (Fig. 6). In some cases, such as CUGN L67Mar salinity (Fig. 6b), the error of DA exp 1.1 is highest and
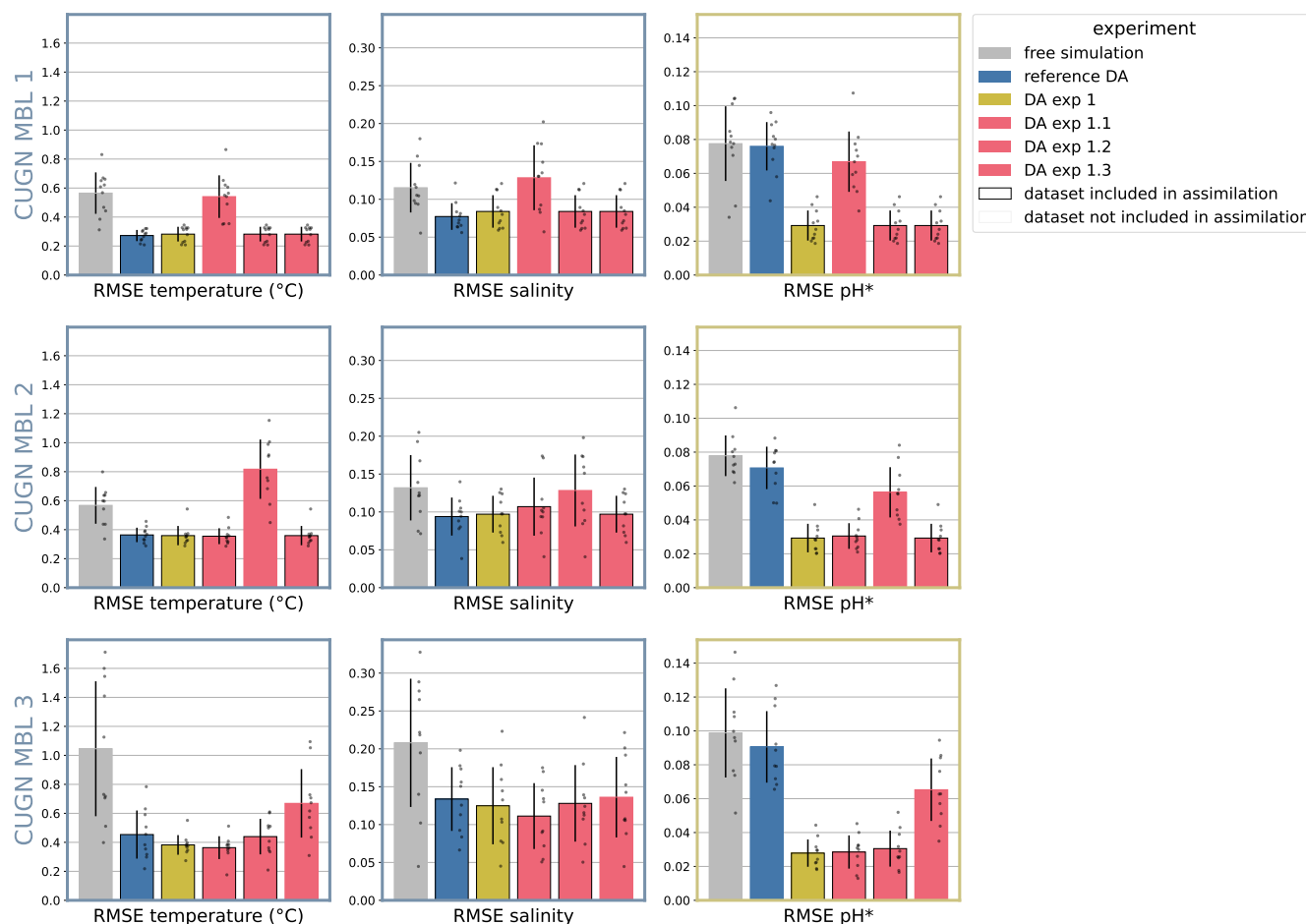
**Figure 6. RMSE in cross-validation experiments:** RMSE quantifying the model-observation misfit for the three CUGN deployments along Line 67, subsets of the CUGN data, used in the cross-validation experiments in Section 3.5. Each panel represents a dataset, the bars represent model experiments: small black dots display the RMSE of a simulation for that dataset, for each cycle which includes data for the dataset. The height of the bar displays the average cycle RMSE, the black line on each bar shows mean ± standard deviation of the cycle RMSE. Black outlines of bars indicate that the dataset was assimilated in the experiment. The cross-validation experiments DA exp 1.1, 1.2, and 1.3 withheld CUGN L67Mar, L67Jul, and L67Oct, respectively.

even exceeds that of the free simulation. This suggests that in some cases, the advection of increments may cause model-data discrepancies downstream. However, for most datasets, the cross-validation experiments perform better than no assimilation
375 of CUGN data. A clear example is the pH RMSE for CUGN L67Mar (Fig. 6c), where the free simulation and reference DA show the highest error (both not assimilating CUGN pH data), while the experiments assimilating CUGN L67Mar data (DA exp 1, 1.2, and 1.3) have the lowest error values. The cross-validation experiment DA exp 1.1 is in between the two results, and benefits from the assimilation of the surrounding CUGN data.

A closer look at pH estimates from cross-validation experiment DA exp 1.1 along CUGN L67Mar reveals that assimilation

380    of neighboring CUGN data positively affects deep pH, largely removing the positive bias seen in free simulation and reference

DA (Fig. 4p-r). However, small patches of high model-data misfit at intermediate depths (≈100 m) indicate that advection of

pH increments from other parts of the CUGN line can cause local increases in error.

In summary, the cross-validation experiments demonstrate that glider DA of pH data typically positively affects pH estimates

near the observation network, but advection of assimilative increments can negatively affect downstream estimates. We examine

385    this interesting point further in Section 4.

### 3.6 Assimilation of oxygen data

Changes in pH and dissolved oxygen are closely related because both are affected by important biogeochemical processes

such as primary production, plankton respiration, and the decomposition of organic matter. These pathways are included in

the NEMUCSC model; however, despite this connection, the assimilation of pH data leads only to minor changes in oxygen

390    in our previous experiments (Fig. 3h, n). Two main factors in the setup of our assimilation system contribute to this limited

impact: (1) The ROMS DA system sets covariance terms between different variables to zero. (2) Our DA configuration favors

corrections to observed variables (such as alkalinity) over corrections to unobserved variables (such as the nutrient variables).

Thus, the assimilation of pH data mainly modifies model alkalinity and DIC and because changes in these two variables do not

directly affect oxygen, the effect of pH DA on oxygen is low. We discuss these factors in more detail in Section 4.

395    To address this limitation, our final set of experiments includes oxygen observations in the assimilation. DA exp 3 assimilates

the reference data along with CUGN pH, alkalinity, and oxygen (Table 2). This additional assimilation of oxygen leads to

marked improvement in oxygen estimates for the CUGN data compared to DA exp 1 (Fig. 3h). It also results in a minor

increase in error for most other assimilated variables (Fig. 3a-g, i). This increase is consistent with the inclusion of more

observations in the DA, which downweights other observations in the 4D-Var state estimation procedure. Importantly, this

400    result does not imply that a better fit to oxygen observations necessarily creates worse estimates for temperature or other

variables.

Notably, DA exp 3 creates better oxygen estimates than the hybrid ESPER-based estimates. This effect mirrors the results

for pH estimates, where direct assimilation in DA exp 1 outperformed the hybrid estimates based on temperature and salinity.

Building on the improved oxygen estimates from DA exp 3, we create an additional hybrid estimate. This new estimate is based

405    on DA exp 3 and ESPER, including oxygen as an input to ESPER alongside temperature and salinity. The resulting hybrid

model produces the best pH estimates and generally outperforms the previously introduced hybrid models, which are based on

temperature and salinity alone.

In summary, the addition of dissolved oxygen observations in the assimilation yields improved oxygen estimates without a

substantial negative effect on other observed variables. Furthermore, the inclusion of assimilated oxygen in hybrid estimates

410    improves their performance compared to those based on temperature and salinity alone. However, it is important to note that

this approach requires a DA system that includes a biogeochemical model with oxygen.

## 4 Discussion

Our study pursued three main goals related to improving carbonate system and oxygen state estimates through data assimilation. First, we examined the effect of the joint assimilation of carbonate system variables and oxygen data with physical and chlorophyll observations that are traditionally assimilated. Our results demonstrate that this joint assimilation successfully improves estimates of pH and oxygen while largely maintaining the quality of physical and chlorophyll estimates. The additional assimilation of glider-based pH and alkalinity data along with oxygen observations achieved substantial reductions in pH and oxygen error in comparison to the reference DA experiment which only assimilated observations of temperature, salinity, sea level anomaly, and chlorophyll. For the reference observations – consisting of satellite-, float-, and glider-based temperature, salinity, sea level anomaly, and chlorophyll – the new data added to the DA system leads only to a minor degradation of the fit in our experiments. This effect is likely due to a down-weighting of the reference observations in the 4D-Var state estimation procedure simply due to the presence of more observations.

Second, we investigated how the assimilation of different data subsets influences the model state estimates. In our configuration, the traditional physical-chlorophyll DA, represented by our reference DA experiment, has minimal impact on pH and oxygen estimates, even when assimilating temperature and salinity data from the same observation platforms that measured pH and oxygen. Similarly, pH DA shows limited influence on oxygen estimates. This limited impact across variables likely stems from two key factors in our DA implementation: the ROMS DA system assumes zero covariance between different variables, and our configuration, in particular the background error specification (see Section 2.5), favors increments to directly observed variables over unobserved ones. As a result, our current DA setup allows larger increments to pH and oxygen directly but not unobserved variables that have pathways with feedback to these variables, such as nitrate, a main driver of primary production that modifies both oxygen and pH.

Third, we evaluated whether combining output from the dynamical ROMS model with the statistical ESPER model can produce improved carbonate system state estimates. Our results show that these hybrid estimates based on ESPER using ROMS temperature and salinity as inputs often produce better results for pH, alkalinity, DIC, and oxygen than the biogeochemical model without assimilating carbonate system variables or oxygen. Importantly, the hybrid estimates benefit from physical DA through improved temperature and salinity estimates and do not require a biogeochemical model. This result suggests that existing physical ocean models and physical DA systems can be used to obtain good carbonate system estimates without implementing complex biogeochemical models.

In this study, we are using super-observations for assimilation while employing the original, non-aggregated observations for visualization and error metrics. This approach proves particularly beneficial for dense glider observations, where multiple measurements often occur within the same model grid cell and time step. Using the original observations for RMSE calculations (Fig. 3, 6) provides a more stringent test of model performance and better represents the true observational density in our visualizations. Interestingly, the model experiments often show a better fit to the original observations than to super-observations. The improved fit is due to the structure of the model grid and vertical differences in the model-observation misfit: the model has terrain-following coordinates, with a higher resolution near the surface, so that deeper grid cells are larger. As

a result, on average, fewer original observations are combined into super-observations near the surface, shifting the proportion of observations upward. Because the model estimates typically have a higher error near the surface, the increased proportion of near-surface super-observations creates a higher mean error when using super-observations.

A significant practical constraint in our pH DA implementation is the requirement for joint assimilation with alkalinity data (see Section 2.6). Since total alkalinity is not routinely measured by autonomous platforms, the assimilation relies on statistical estimates. This dependence on estimated alkalinity data represents a potential limitation of our approach in cases when the alkalinity estimates may not be accurate. For example, this DA approach has limited use in cases where the alkalinity has been altered (e.g., as part of a marine carbon dioxide removal (mCDR) experiment (Boyd et al., 2023; Fennel et al., 2023)). Other instances of unreliable estimates from global algorithms like ESPER may include regions near river mouths with significant organic matter input from terrestrial sources or following large denitrification events at the end of a seasonal phytoplankton bloom. In these latter cases, regional statistical models, specifically trained for their area of interest, may be able to produce more accurate local estimates. If there has been no alteration and statistical estimates of alkalinity are reliable, our results suggest that the statistical estimates are sufficiently accurate for effective pH state estimation.

The use of statistical models for both data generation and estimation in our study warrants careful consideration. In our current setup, we assimilate ESPER-based pH and alkalinity data for the CUGN lines, as well as ESPER-based alkalinity for the pH-sensor glider line. Simultaneously, we use ESPER for our hybrid estimates, creating a close relationship between data and the hybrid estimates. This relationship is most obvious in the hybrid estimates from the DA experiment assimilating CUGN pH, alkalinity, and oxygen data in addition to the reference data (ESPER estimates based on DA exp 3, see Section 3.6): the CUGN pH data was generated using ESPER based on glider measurements of temperature, salinity and oxygen. The hybrid estimates were generated using ESPER based on model estimates of temperature, salinity, and oxygen from an experiment assimilating the same glider measurements. Hence, the good fit of the hybrid estimates to pH data (see Fig. 3 g) is less a result of the quality of the ESPER estimates and more a result of the DA system's ability to move model estimates closer to the glider observations. Despite this limitation, hybrid estimates also show a good fit to measured pH data in our experiments (Fig. 3 m) and even the ESPER estimates based on the free simulation (and only temperature and salinity input) outperform pH model estimates from simulations that did not assimilate pH data. Hence, dependent on the setup and scope of the study, we consider hybrid estimates as valid alternatives to complex biogeochemical models.

Looking ahead, several important challenges emerge from this study that warrant further investigation. The temporal scope of our analysis, limited to 2019, leaves open questions about the long-term effectiveness of the DA system. Here, we are planning to extend the DA beyond 2019, complemented by validation against data not included in the DA system, for example, from the BGC-Argo program (Bittig et al., 2019) and recent high-resolution $pCO_2$ products for the CCS (Sharp et al., 2022). We further aim to address current methodological limitations through modification of background error values in the 4D-Var system or the implementation of ensemble-based DA (e.g., using the framework presented in Mattern and Edwards (2023)) to better account for cross-variable covariances.

By demonstrating the feasibility of joint DA of physical, chlorophyll, and carbonate system variables and hybrid estimation approaches, this study provides practical pathways for improving carbonate system state estimates in dynamic coastal environ-

ments. These methodological advances support broader efforts to quantify regional carbon budgets and air-sea $CO_2$ exchange, contributing to our understanding of how coastal systems respond to and influence global carbon cycling. The framework developed here is particularly relevant as ocean observing systems expand and the need for accurate coastal carbon cycle monitoring grows more urgent in the face of climate change and ocean-based mitigation efforts like mCDR.

## 5 Conclusions

In this study, we examined the effects of integrating glider-based pH, alkalinity, and oxygen data into a joint physical-biogeochemical DA system. In our reference DA configuration, the assimilation of physical and chlorophyll data has little effect on the model state estimates for pH or oxygen. However, the joint assimilation of pH, alkalinity, and oxygen data substantially improves carbonate system estimates without degrading the quality of physical and chlorophyll estimates. Cross-validation experiments reveal that pH data assimilation typically improves estimates near the observation network, although downstream advection of increments can occasionally degrade results. A hybrid approach, which combines the output of the dynamical ROMS model with ESPER statistical estimates of pH and oxygen, produces good carbonate system estimates that benefit from physical data assimilation while eliminating the need for a biogeochemical model. While this study is limited to 2019, it is the basis for extending the DA to the following years and examining the continued impact on the carbonate system and ocean acidification state estimates in a highly dynamic coastal ecosystem.

*Data availability.* The data used for this study is publicly available; data sources are listed in Table 1.

*Author contributions.* Jann Paul Mattern: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation; Yuichiro Takeshita: Conceptualization, Funding Acquisition, Writing – Original Draft Preparation; Carlos Rocha: Conceptualization, Methodology; Christopher A. Edwards: Conceptualization, Methodology, Funding Acquisition, Writing – Original Draft Preparation

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The authors thank Joseph Warren, Jacki Long, Brent Jones, and the Instrument Development Group at the Scripps Institution of Oceanography for their assistance with Monterey Bay glider operations and data quality control. JPM and CAE gratefully acknowledge funding support from the National Oceanic and Atmospheric Administration's Ocean Acidification Program (grants NA19OAR0170357 and NA24OARX017G0021-T1-01) and the Simons Collaboration on Computational Biogeochemical Modeling of Ma-

rine Ecosystems (award 549949). We thank the David and Lucile Packard Foundation for supporting YT and the Monterey Bay glider operations.

# References

Alin, S. R., Feely, R. A., Dickson, A. G., Hernández-Ayón, J. M., Juranek, L. W., Ohman, M. D., and Goericke, R.: Robust empirical
   relationships for estimating the carbonate system in the southern California Current System and application to CalCOFI hydrographic
   cruise data (2005–2011), Journal of Geophysical Research: Oceans, 117, https://doi.org/10.1029/2011JC007511, 2012.

Bakker, D. C., Alin, S. R., Becker, M., Bittig, H. C., Castaño-Primo, R., Feely, R. A., Gkritzalis, T., Kadono, K., Kozyr, A., Lauvset, S. K.,
   Metzl, N., Munro, D. R., Nakaoka, S.-i., Nojiri, Y., O'Brien, K. M., Olsen, A., Pfeil, B., Pierrot, D., Steinhoff, T., Sullivan, K. F., Sutton,
   A. J., Sweeney, C., Tilbrook, B., Wada, C., Wanninkhof, R., Willstrand Wranne, A., Akl, J., Apelthun, L. B., Bates, N., Beatty, C. M.,
   Burger, E. F., Cai, W.-J., Cosca, C. E., Corredor, J. E., Cronin, M., Cross, J. N., De Carlo, E. H., DeGrandpre, M. D., Emerson, S.,
   Enright, M. P., Enyo, K., Evans, W., Frangoulis, C., Fransson, A., García-Ibáñez, M. I., Gehrung, M., Giannoudi, L., Glockzin, M., Hales,
   B., Howden, S. D., Hunt, C. W., Ibánhez, J. S., Jones, S. D., Kamb, L., Körtzinger, A., Landa, C. S., Landschützer, P., Lefèvre, N.,
   Lo Monaco, C., Macovei, V. A., Maenner Jones, S., Meinig, C., Millero, F. J., Monacci, N. M., Mordy, C., Morell, J. M., Murata, A.,
   Musielewicz, S., Neill, C., Newberger, T., Nomura, D., Ohman, M., Ono, T., Passmore, A., Petersen, W., Petihakis, G., Perivoliotis, L.,
   Plueddemann, A. J., Rehder, G., Reynaud, T., Rodriguez, C., Ross, A. C., Rutgersson, A., Sabine, C. L., Salisbury, J. E., Schlitzer, R.,
   Send, U., Skjelvan, I., Stamataki, N., Sutherland, S. C., Sweeney, C., Tadokoro, K., Tanhua, T., Telszewski, M., Trull, T., Vandemark, D.,
   van Ooijen, E., Voynova, Y. G., Wang, H., Weller, R. A., Whitehead, C., and Wilson, D.: Surface Ocean CO2 Atlas Database Version 2022
   (SOCATv2022) (NCEI Accession 0253659), https://doi.org/10.25921/1h9f-nb73, 2022.

Bauer, J. E., Cai, W. J., Raymond, P. A., Bianchi, T. S., Hopkinson, C. S., and Regnier, P. A.: The changing carbon cycle of the coastal ocean,
   Nature, 504, 61–70, https://doi.org/10.1038/nature12857, 2013.

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., Körtzinger, A., and Gattuso, J.-P.: An Alternative to Static
   Climatologies: Robust Estimation of Open Ocean CO2 Variables and Nutrient Concentrations From T, S, and O2 Data Using Bayesian
   Neural Networks, Frontiers in Marine Science, 5, 1–29, https://doi.org/10.3389/fmars.2018.00328, 2018.

Bittig, H. C., Maurer, T. L., Plant, J. N., Schmechtig, C., Wong, A. P. S., Claustre, H., Trull, T. W., Udaya Bhaskar, T. V. S., Boss, E.,
   Dall'Olmo, G., Organelli, E., Poteau, A., Johnson, K. S., Hanstein, C., Leymarie, E., Le Reste, S., Riser, S. C., Rupan, A. R., Taillandier,
   V., Thierry, V., and Xing, X.: A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage, Frontiers in Marine Science, 6,
   https://doi.org/10.3389/fmars.2019.00502, 2019.

Boyd, P., Claustre, H., Legendre, L., Gattuso, J.-P., and Le Traon, P.-Y.: Operational Monitoring of Open-Ocean Car-
   bon Dioxide Removal Deployments: Detection, Attribution, and Determination of Side Effects, Oceanography, 36, 1–10,
   https://doi.org/10.5670/oceanog.2023.s1.2, 2023.

Carroll, D., Menemenlis, D., Adkins, J. F., Bowman, K. W., Brix, H., Dutkiewicz, S., Fenty, I., Gierach, M. M., Hill, C., Jahn, O., Land-
   schützer, P., Lauderdale, J. M., Liu, J., Manizza, M., Naviaux, J. D., Rödenbeck, C., Schimel, D. S., Van der Stocken, T., and Zhang,
   H.: The ECCO-Darwin Data-Assimilative Global Ocean Biogeochemistry Model: Estimates of Seasonal to Multidecadal Surface Ocean
   pCO2 and Air-Sea CO2 Flux, Journal of Advances in Modeling Earth Systems, 12, 1–28, https://doi.org/10.1029/2019MS001888, 2020.

Carter, B. R., Bittig, H. C., Fassbender, A. J., Sharp, J. D., Takeshita, Y., Xu, Y. Y., Álvarez, M., Wanninkhof, R., Feely, R. A., and Barbero,
   L.: New and updated global empirical seawater property estimation routines, Limnology and Oceanography: Methods, 19, 785–809,
   https://doi.org/10.1002/lom3.10461, 2021.

Checkley, D. M. and Barth, J. a.: Patterns and processes in the California Current System, Progress in Oceanography, 83, 49–64,
   https://doi.org/10.1016/j.pocean.2009.07.028, 2009.

545 Cheresh, J. and Fiechter, J.: Physical and Biogeochemical Drivers of Alongshore pH and Oxygen Variability in the California Current System, Geophysical Research Letters, 47, 1–9, https://doi.org/10.1029/2020GL089553, 2020.

Deutsch, C., Frenzel, H., Mcwilliams, J. C., Renault, L., Kessouri, F., Howard, E., Liang, J.-h., Bianchi, D., and Yang, S.: Progress in Oceanography Biogeochemical variability in the California Current System, Progress in Oceanography, 196, 102 565, https://doi.org/10.1016/j.pocean.2021.102565, 2021.

550 Doyle, J. D., Jiang, Q., Chao, Y., and Farrara, J.: High-resolution real-time modeling of the marine atmospheric boundary layer in support of the AOSN-II field campaign, Deep-Sea Research Part II: Topical Studies in Oceanography, 56, 87–99, https://doi.org/10.1016/j.dsr2.2008.08.009, 2009.

Edwards, C. A., Moore, A. M., Hoteit, I., and Cornuelle, B. D.: Regional Ocean Data Assimilation, Annual Review of Marine Science, 7, 21–42, https://doi.org/10.1146/annurev-marine-010814-015821, 2015.

555 Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C. A., Ford, D., Friedrichs, M. A., Gregoire, M., Jones, E., Kim, H. C., Lamouroux, J., Murtugudde, R., and Perruche, C.: Advancing marine biogeochemical and ecosystem reanalyses and forecasts as tools for monitoring and managing ecosystem health, Frontiers in Marine Science, 6, 1–9, https://doi.org/10.3389/fmars.2019.00089, 2019.

Fennel, K., Mattern, J. P., Doney, S. C., Bopp, L., Moore, A. M., Wang, B., and Yu, L.: Ocean biogeochemical modelling, Nature Reviews
560 Methods Primers, 2, 76, https://doi.org/10.1038/s43586-022-00154-2, 2022.

Fennel, K., Long, M. C., Algar, C. K., Carter, B., Keller, D., Laurent, A., Mattern, J. P., Musgrave, R., Oschlies, A., Ostiguy, J., Palter, J., and Whitt, D. B.: Modelling considerations for research on ocean alkalinity enhancement (OAE), State of the Planet Discussions, pp. 1–47, https://doi.org/10.5194/sp-2-oae2023-9-2023, 2023.

Fiechter, J., Edwards, C. A., and Moore, A. M.: Wind , Circulation , and Topographic Effects on Alongshore Phytoplankton Variability in
565 the California Current, Geophysical Research Letters, pp. 3238–3245, https://doi.org/10.1002/2017GL076839, 2018.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Barbero, L., Bates, N. R., Becker, M., Bellouin, N., Decharme, B., Bopp, L., Brasika, I. B. M., Cadule, P., Chamberlain, M. A., Chandra, N., Chau, T.-T.-T., Chevallier, F., Chini, L. P., Cronin, M., Dou, X., Enyo, K., Evans, W., Falk, S., Feely, R. A., Feng, L., Ford, D. J.,
570 Gasser, T., Ghattas, J., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Joos, F., Kato, E., Keeling, R. F., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot, N., McGuire, P. C., McKinley, G. A., Meyer, G., Morgan, E. J., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K. M., Olsen, A., Omar, A. M., Ono, T., Paulsen, M., Pierrot, D., Pocock, K., Poulter, B., Powis, C. M., Rehder, G., Resplandy, L.,
575 Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Séférian, R., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tans, P. P., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., van Ooijen, E., Wanninkhof, R., Watanabe, M., Wimart-Rousseau, C., Yang, D., Yang, X., Yuan, W., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.: Global Carbon Budget 2023, Earth System Science Data, 15, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023, 2023.

Garcia, H. E., Bouchard, C., Cross, S. L., Paver, C. R., Reagan, J. R., Boyer, T. P., Locarnini, R. A., Mishonov, A. V., Baranova, O. K.,
580 Seidov, D., Wang, Z., and Dukhovskoy, D.: World Ocean Atlas 2023, Volume 4: Dissolved Inorganic Nutrients (Phosphate, Nitrate, and Silicate), NOAA Atlas NESDIS 92, 4, 79pp, https://doi.org/10.25923/39qw-7j08, 2024.

Green, R. A., Edwards, C. A., Rafter, P. A., and Hain, M. P.: Assessing the Effectiveness and Detectability of Ocean Alkalinity Enhancement in the California Current System using Regional Ocean Model Simulations , Earth's Future, in prep.

Haidvogel, D. B., Arango, H. G., Budgell, W. P., Cornuelle, B. D., Curchitser, E. N., Di Lorenzo, E., Fennel, K., Geyer, W. R., Hermann, A. J., Lanerolle, L., Levin, J., McWilliams, J. C., Miller, A. J., Moore, A. M., Powell, T. M., Shchepetkin, A. F., Sherwood, C. R., Signell, R. P., Warner, J. C., and Wilkin, J.: Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System, Journal of Computational Physics, 227, 3595–3624, https://doi.org/10.1016/j.jcp.2007.06.016, 2008.

Hemming, M. P., Kaiser, J., Heywood, K. J., Bakker, D. C., Boutin, J., Shitashima, K., Lee, G., Legge, O., and Onken, R.: Measuring pH variability using an experimental sensor on an underwater glider, Ocean Science, 13, 427–442, https://doi.org/10.5194/os-13-427-2017, 2017.

Jackson, T., Sathyendranath, S., and Platt, T.: An Exact Solution For Modeling Photoacclimation of the Carbon-to-Chlorophyll Ratio in Phytoplankton, Frontiers in Marine Science, 4, 1–10, https://doi.org/10.3389/fmars.2017.00283, 2017.

Jean-Michel, L., Eric, G., Romain, B.-B., Gilles, G., Angélique, M., Marie, D., Clément, B., Mathieu, H., Olivier, L. G., Charly, R., Tony, C., Charles-Emmanuel, T., Florent, G., Giovanni, R., Mounir, B., Yann, D., and Pierre-Yves, L. T.: The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis, Frontiers in Earth Science, 9, 1–27, https://doi.org/10.3389/feart.2021.698876, 2021.

Johnson, K. S., Jannasch, H. W., Coletti, L. J., Elrod, V. A., Martz, T. R., Takeshita, Y., Carlson, R. J., and Connery, J. G.: Deep-Sea DuraFET: A Pressure Tolerant pH Sensor Designed for Global Sensor Networks, Analytical Chemistry, 88, 3249–3256, https://doi.org/10.1021/acs.analchem.5b04653, 2016.

Kishi, M. J., Kashiwai, M., Ware, D. M., Megrey, B. A., Eslinger, D. L., Werner, F. E., Noguchi-Aita, M., Azumaya, T., Fujii, M., Hashimoto, S., Huang, D., Iizumi, H., Ishida, Y., Kang, S., Kantakov, G. A., Kim, H.-c., Komatsu, K., Navrotsky, V. V., Smith, S. L., Tadokoro, K., Tsuda, A., Yamamura, O., Yamanaka, Y., Yokouchi, K., Yoshie, N., Zhang, J., Zuenko, Y. I., and Zvalin-sky, V. I.: NEMURO—a lower trophic level model for the North Pacific marine ecosystem, Ecological Modelling, 202, 12–25, https://doi.org/10.1016/j.ecolmodel.2006.08.021, 2007.

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, 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.

Land, P. E., Findlay, H. S., Shutler, J. D., Piolle, J. F., Sims, R., Green, H., Kitidis, V., Polukhin, A., and Pipko, I. I.: OceanSODA-MDB: a standardised surface ocean carbonate system dataset for model-data intercomparisons, Earth System Science Data, 15, 921–947, https://doi.org/10.5194/essd-15-921-2023, 2023.

Martin, M. J., Hoteit, I., Bertino, L., and Moore, A. M.: Data assimilation schemes for ocean forecasting: state of the art, State of the Planet Discussions, 2024, 1–16, https://doi.org/10.5194/sp-2024-20, 2024.

Mathis, M., Lacroix, F., Hagemann, S., Nielsen, D. M., Ilyina, T., and Schrum, C.: Enhanced CO2 uptake of the coastal ocean is dominated by biological carbon fixation, Nature Climate Change, 14, 373–379, https://doi.org/10.1038/s41558-024-01956-w, 2024.

Mattern, J. P. and Edwards, C. A.: Ensemble optimal interpolation for adjoint-free biogeochemical data assimilation, PLOS ONE, 18, e0291039, https://doi.org/10.1371/journal.pone.0291039, 2023.

Mattern, J. P., Song, H., Edwards, C. A., Moore, A. M., and Fiechter, J.: Data assimilation of physical and chlorophyll a observations in the California Current System using two biogeochemical models, Ocean Modelling, 109, 55–71, https://doi.org/10.1016/j.ocemod.2016.12.002, 2017.

Mattern, J. P., Edwards, C. A., and Moore, A. M.: Improving Variational Data Assimilation through Background and Observation Error Adjustments, Monthly Weather Review, 146, 485–501, https://doi.org/10.1175/MWR-D-17-0263.1, 2018.

Olsen, A., Lange, N., Key, R. M., Tanhua, T., Bittig, H. C., Kozyr, A., Álvarez, M., Azetsu-Scott, K., Becker, S., Brown, P. J., Carter, B. R., Cotrim Da Cunha, L., Feely, R. A., Van Heuven, S., Hoppema, M., Ishii, M., Jeansson, E., Jutterström, S., Landa, C. S., Lauvset, S. K., Michaelis, P., Murata, A., Pérez, F. F., Pfeil, B., Schirnick, C., Steinfeldt, R., Suzuki, T., Tilbrook, B., Velo, A., Wanninkhof, R., and Woosley, R. J.: An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2020, Earth System Science Data, 12, 3653–3678, https://doi.org/10.5194/essd-12-3653-2020, 2020.

Ren, A. S., Rudnick, D. L., and Twombly, A.: Drift Characteristics of Sea-Bird Dissolved Oxygen Optode Sensors, Journal of Atmospheric and Oceanic Technology, 40, 1457–1468, https://doi.org/10.1175/JTECH-D-22-0103.1, 2023.

Roobaert, A., Laruelle, G. G., Landschützer, P., Gruber, N., Chou, L., and Regnier, P.: The Spatiotemporal Dynamics of the Sources and Sinks of $CO_2$ in the Global Coastal Ocean, Global Biogeochemical Cycles, 33, 1693–1714, https://doi.org/10.1029/2019GB006239, 2019.

Rudnick, D. L.: Ocean Research Enabled by Underwater Gliders, Annual Review of Marine Science, 8, 519–541, https://doi.org/10.1146/annurev-marine-122414-033913, 2016.

Rudnick, D. L., Zaba, K. D., Todd, R. E., and Davis, R. E.: A climatology of the California Current System from a network of underwater gliders, Progress in Oceanography, 154, 64–106, https://doi.org/10.1016/j.pocean.2017.03.002, 2017.

Saba, G. K., Wright-Fairbanks, E., Chen, B., Cai, W. J., Barnard, A. H., Jones, C. P., Branham, C. W., Wang, K., and Miles, T.: The Development and Validation of a Profiling Glider Deep ISFET-Based pH Sensor for High Resolution Observations of Coastal and Ocean Acidification, Frontiers in Marine Science, 6, 1–17, https://doi.org/10.3389/fmars.2019.00664, 2019.

Sharp, J. D., Fassbender, A. J., Carter, B. R., Lavin, P. D., and Sutton, A. J.: A monthly surface pCO2 product for the California Current Large Marine Ecosystem, Earth System Science Data, 14, 2081–2108, https://doi.org/10.5194/essd-14-2081-2022, 2022.

Song, H., Edwards, C. A., Moore, A. M., and Fiechter, J.: Incremental four-dimensional variational data assimilation of positive-definite oceanic variables using a logarithm transformation, Ocean Modelling, 54-55, 1–17, https://doi.org/10.1016/j.ocemod.2012.06.001, 2012.

Takeshita, Y., Jones, B. D., Johnson, K. S., Chavez, F. P., Rudnick, D. L., Blum, M., Conner, K., Jensen, S., Long, J. S., Maughan, T., Mertz, K. L., Sherman, J. T., and Warren, J. K.: Accurate ph and o2 measurements from spray underwater gliders, Journal of Atmospheric and Oceanic Technology, 38, 181–195, https://doi.org/10.1175/JTECH-D-20-0095.1, 2021.

Turner, K. E., Smith, D. M., Katavouta, A., and Williams, R. G.: Reconstructing ocean carbon storage with CMIP6 models and synthetic Argo observations, Biogeosciences, 20, 1671–1690, https://doi.org/10.5194/bg-20-1671-2023, 2023.

Veneziani, M., Edwards, C. A., Doyle, J. D., and Foley, D.: A central California coastal ocean modeling study: 1. Forward model and the influence of realistic versus climatological forcing, Journal of Geophysical Research, 114, 1–16, https://doi.org/10.1029/2008JC004774, 2009.

Verdy, A. and Mazloff, M. R.: A data assimilating model for estimating Southern Ocean biogeochemistry, Journal of Geophysical Research: Oceans, 122, 6968–6988, https://doi.org/10.1002/2016JC012650, 2017.

Visinelli, L., Masina, S., Vichi, M., Storto, A., and Lovato, T.: Impacts of data assimilation on the global ocean carbonate system, Journal of Marine Systems, 158, 106–119, https://doi.org/10.1016/j.jmarsys.2016.02.011, 2016.

655  Wolfe, W. H., Martz, T. R., Dickson, A. G., Goericke, R., and Ohman, M. D.: A 37-year record of ocean acidification in the Southern
California current, Communications Earth and Environment, 4, 1–7, https://doi.org/10.1038/s43247-023-01065-0, 2023.