

# Hybrid machine learning data assimilation for marine biogeochemistry

Ieuan Higgs<sup>1,2</sup>, Ross Bannister<sup>1,2</sup>, Jozef Skákala<sup>2,3</sup>, Alberto Carrassi<sup>1,4</sup>, and Stefano Ciavatta<sup>5</sup>

<sup>1</sup>Department of Meteorology, University of Reading, UK

<sup>2</sup>National Centre for Earth Observation, UK

<sup>3</sup>Plymouth Marine Laboratory, UK

<sup>4</sup>Department of Physics and Astronomy “Augusto Righi”, University of Bologna, IT

<sup>5</sup>Mercator Ocean International, FR

September 22, 2025

[Corresponding Author: Ieuan Higgs \(i.c.higgs@reading.ac.uk\)](mailto:i.c.higgs@reading.ac.uk)

## Abstract

Marine biogeochemistry models are critical for forecasting, as well as estimating ecosystem responses to climate change and human activities. Data assimilation (DA) improves these models by aligning them with real-world observations, but marine biogeochemistry DA faces challenges due to model complexity, strong nonlinearity, and sparse, uncertain observations. Existing DA methods applied to marine biogeochemistry struggle to update unobserved variables effectively, while ensemble-based methods are computationally too expensive for high-complexity marine biogeochemistry models. This study demonstrates how machine learning (ML) can improve marine biogeochemistry DA by learning statistical relationships between observed and unobserved variables. We integrate ML-driven balancing schemes into a 1D prototype of a system used to forecast marine biogeochemistry in the North-West European Shelf seas. ML is applied to ~~predict~~ estimate (i) state-dependent correlations from free-run ensembles and (ii), in an “end-to-end” fashion, analysis increments from an Ensemble Kalman Filter. Our results show that ML significantly enhances updates for previously not-updated variables when compared to univariate schemes akin to those used operationally. Furthermore, ML models exhibit moderate transferability to new locations, a crucial step toward scaling these methods to 3D operational systems. We conclude that ML offers a clear pathway to overcome current computational bottlenecks in marine biogeochemistry DA and that refining transferability, optimizing training data sampling, and evaluating scalability for large-scale marine forecasting, should be future research priorities.

## 1 Introduction

Marine biogeochemistry (BGC) modelling is an essential tool for understanding global marine elemental cycles (e.g., for carbon and nitrogen), as well as for understanding the response of marine ecosystems to a range of human and climate pressures (Heinze and Gehlen, 2013; Ford et al., 2018; Fennel et al., 2022). These pressures include ocean acidification, marine heat waves, and nutrient pollution ~~and which can~~ lead to a range of consequences, such as deoxygenation, toxic algal blooms and biodiversity loss (Doney et al., 2009; Smith and Schindler, 2009; Schmidtko et al., 2017; Frölicher and Laufkötter, 2018; Fennel and Testa, 2019; Gobler, 2020). Marine BGC modelling could then support management, policy and planning across a wide range of temporal scales. Marine BGC models are often constrained by the available observations through data assimilation (DA) (Ford et al., 2018; Fennel et al., 2019), providing both multi-decadal reanalyses of past ecosystem trends and variability, as well as short-term operational forecasts (on the scale up to 5-10 days). Such operational forecasts are run by marine forecasting centres in many countries, e.g., by the Copernicus Marine Service in Europe covering the global ocean and all the major European seas (Le Traon et al., 2019).

However, marine BGC DA faces multiple specific challenges (Dowd et al., 2014; Ford et al., 2018; Fennel et al., 2019), compared to assimilation of ocean physics observations in marine models. Marine BGC models are typically more complex than physical models (a pelagic model can have tens of

48 variables and hundreds of parameters), they are highly non-linear and relatively poorly constrained  
49 (e.g., having highly uncertain parameters) when compared to ocean physics models. Furthermore,  
50 marine BGC observations are even fewer, sparser, and more uncertain than physics observations. This  
51 brings several specific challenges for marine BGC DA, one of those being the need for multivariate  
52 DA, where a large portion of the marine BGC model state variables is updated by observations of  
53 only a small fraction of the model variables. In the context of operational marine BGC forecasting,  
54 these observations are typically satellite ocean colour-derived chlorophyll (Fennel et al., 2019; Groom  
55 et al., 2019), with assimilation of BGC-Argo observations (including chlorophyll, nitrate and oxygen)  
56 in open ocean waters recently implemented in state-of-the-art operational systems (Cossarini et al.,  
57 2019; Teruzzi et al., 2021). Other products are assimilated in reanalyses or research and development  
58 (R&D) versions of the operational systems, such as optical variables (Shulman et al., 2013; Ciavatta  
59 et al., 2014; Jones et al., 2016; Gregg and Rousseaux, 2017; Skakala et al., 2020) ~~and size-class~~  
60 ~~chlorophyll (Ciavatta et al., 2018, 2019; Skákala et al., 2018; Pradhan et al., 2020) and other~~, as  
61 well as types of in situ data, such as chlorophyll, oxygen and nutrients from gliders (Skákala et al.,  
62 2021) ~~are assimilated in reanalyses, or research and development (R&D) versions of the operational~~  
63 ~~systems~~. For a broader range of marine BGC DA work beyond operational applications, see many  
64 other references, e.g., Simon and Bertino (2012); Shulman et al. (2013); Gehlen et al. (2015); Simon  
65 et al. (2015).

66 Different DA systems are used across marine BGC forecasting centres, including variational (Ford  
67 et al., 2012; Song et al., 2016; Skákala et al., 2018; Coppini et al., 2021), Singular Evolutive Extended  
68 Kalman filter (SEEK) (Gutknecht et al., 2019; Ciliberti et al., 2021) and Ensemble Kalman Filter  
69 (EnKF) (Bertino et al., 2021) -based methods. Although ensemble methods (e.g., EnKF) are appeal-  
70 ing for their capability to provide uncertainty quantification and cross-covariances, the more complex  
71 marine BGC models such as the European Regional Seas Ecosystem Model (ERSEM) (Butenschön  
72 et al., 2016) or the Biogeochemical Flux Model (BFM) (Cossarini et al., 2017), currently rely on vari-  
73 ational methods ~~as~~ running a sufficiently large ensemble in the day-to-day operational forecasting  
74 context can be computationally ~~prohibitive~~expensive. Moreover, for such complex models, variational  
75 methods update only a very limited number of unobserved variables, typically using very simple bal-  
76 ancing principles based on the simulated structure and stoichiometry of the phytoplankton community  
77 (Teruzzi et al., 2014; Skákala et al., 2018). We will call such systems with certain approximations  
78 “univariate”, and systems that update (nearly) all model state variables as a direct result of DA  
79 “multivariate”. The multivariate updates can happen in the DA step, through ensemble-informed  
80 background<sup>1</sup> error covariances (as in the EnKF), ~~or~~, through balancing schemes, such as the scheme  
81 of Hemmings et al. (2008) based on nitrogen mass conservation applied to Nutrient-Phytoplankton-  
82 Zooplankton-Detritus models (Hemmings et al., 2008; Ford et al., 2012) ~~), or through the tangent-linear~~  
83 and adjoint models Mattern et al. (2017). However, whenever such multivariate schemes were applied  
84 to highly complex marine BGC models (in reanalyses, or R&D), the improvement on non-observed  
85 variables was typically marginal, with several variables often systematically degraded by DA (e.g., Cia-  
86 vatta et al. (2016, 2018)). This provides a warning on the use of incorrect assumptions in multivariate  
87 balancing schemes or in the EnKFs ~~and~~ the need for better DA and/or ensemble design.

88 The field of machine learning (ML) has developed rapidly during the past few decades, and has  
89 seemingly found function across every level of science and culture, due to the increasing size and  
90 availability of datasets and computational power, together with the continued development of algo-  
91 rithms and theory (Jordan and Mitchell, 2015; Sonnewald et al., 2021). Within Earth sciences, the  
92 flexibility of ML paradigms has allowed its use in a huge variety of applications (Reichstein et al.,  
93 2019), including extensive use in physical ocean modelling (van der Merwe et al., 2007; Nowack et al.,  
94 2018; Kochkov et al., 2021). However, using ML for marine BGC models is comparatively infrequent,  
95 with the most common examples found in parameter estimation (Mattern et al., 2012; Leeds et al.,  
96 2013; Mattern et al., 2014; Schartau et al., 2017). There are only relatively few applications outside  
97 of this domain such as using a statistical emulator to quantify uncertainty (Mattern et al., 2013) and  
98 the prediction of hypoxia in shelf sea environments (Skakala et al., 2023).

99 In this work, we investigate the capability of ML to learn the ~~hidden, non-linear, and complex~~  
100 /flow-dependent relations between BGC variables ~~and to use the~~, before using those learned functions  
101 within a DA scheme or to fully substitute it. Thus, we are not attempting to emulate or improve  
102 BGC models via ML, but instead use ML to improve DA, and specifically to cope with the chal-  
103 lenging problem of propagating information from observed to unobserved variables in single-model  
104 deterministic runs. The main goal of this approach is to introduce multivariate DA into the system,  
105 whilst benefiting from the relatively low computational cost of ML. This study falls within a stream

---

<sup>1</sup>In data assimilation, the terms “forecast” and “background” are often used interchangeably. Strictly, the background refers to the forecast state used as the prior in the assimilation step.

106 of research aimed at building suitable hybrid ML-DA schemes (see Buizza et al., 2022; Cheng et al.,  
107 2023, and references therein), and, to our knowledge, it is the first such attempt in the context of  
108 marine BGC.

109 We first use ML to learn flow-dependent correlations that are needed within a DA update step.  
110 This amounts to a merge of DA and ML, whereby the latter is used to accomplish a task within the  
111 DA process. We demonstrate that such an ML-based multivariate DA is efficient and accurate. As  
112 long as enough suitable data are available for training, ML is able to learn and map complex non-linear  
113 functions for propagating the information from observed to unobserved portions of the system’s state.

114 In a second configuration, instead of merging DA and ML, the former is used to produce a training  
115 dataset from which ML learns the full DA step, in an “end-to-end” fashion (Barth et al., 2020;  
116 Fablet et al., 2021). Here, the ML task is that of DA as a whole, i.e., given the background state  
117 and observations, return the analysis increments to the background for unobserved variables. As  
118 mentioned above, we do not intend substituting/improving the BGM model, and our end-to-end  
119 learning focuses only on learning the instantaneous DA updates while using the given BGC model to  
120 issue the forecasts. Efficient end-to-end learning of the EnKF analysis in chaotic systems has been  
121 recently proven by Bocquet et al. (2024).

122 Specifically, we intend to answer the following questions: (a) Can we make improvements to  
123 the existing univariate scheme by updating a limited set of additional variables with an ML model  
124 to ~~predict-estimate~~ correlations or analysis increments? (b) Can these ML models be extended to  
125 effectively update all unobserved pelagic variables? (c) Is the ML model transferable to a new location  
126 after being trained on some other location?

127 Our work has a potentially important application within the North-West European Shelf (NWES)  
128 operational DA system to which it is tailored. Yet we will discuss its ~~generalization-generalisation~~  
129 to other comparable systems, applied to spatial domains with similar type of marine BGC dynamics.  
130 Based on the transferability of the ML model, we speculate whether it is feasible to use the ML model  
131 trained in 1D on a 3D domain and propose a methodology for doing so.

132 The paper is structured as follows. We first give, in Sect. 2, details on the 1D physical model, the  
133 BGC model and the configuration used. Also, we establish the setups for the DA workflow, describing  
134 the reference univariate scheme (RUS), and the use of the EnKF. Then, in Sect. 3, we outline the two  
135 ML approaches explored in this work. We also give detail on the ML architecture and climatological  
136 statistics. Next, in Sect. 4, we present and discuss our results for: updating nitrate only; updating  
137 the entire set of pelagic BGC surface variables; and testing the transferability of the ML model to a  
138 new location with different BGC behaviour. In Sect. 5, we draw concluding remarks, summarise the  
139 key findings and discuss future work.

## 140 2 Model and data assimilation setups for biogeochemistry

### 141 2.1 Physical model: GOTM

142 The Generalised Ocean Turbulence Model (GOTM) (Bolding and Villarreal, 1999) is a 1D water  
143 column model for studying hydrodynamic and biogeochemical processes when coupled to a biogeo-  
144 chemical model, in marine and limnic waters. It provides a ~~sufficient~~ balance between realism and  
145 computational cost ~~, posing as an ideal candidate for studying new DA schemes in realistic scenarios~~  
146 by using real atmospheric forcing data, relaxation profiles, and coupling at full BGC complexity, while  
147 sacrificing the explicit representation of 3D processes. This makes the system ideal for this work, where  
148 we are primarily interested in the error relationships between different biogeochemical quantities (e.g.,  
149 chlorophyll to nitrate), rather than the spatial error characteristics. GOTM can be used as a stand-  
150 alone model for studying dynamics of boundary layers in natural waters, having hydrodynamic appli-  
151 cations in investigations of air-sea fluxes (Vagle et al., 2010), surface mixed-layer dynamics (Sonntag  
152 and Hense, 2011), dynamics of bottom boundary layers with or without sediment transport (Umlauf  
153 and Burchard, 2011; Falchetti et al., 2010), and estuarine and coastal dynamics (Burchard, 2009).

### 154 2.2 Biogeochemical model: ERSEM

155 ERSEM (Baretta et al., 1995; Butenschön et al., 2016) is a marine biogeochemistry model that sim-  
156 ulates lower trophic levels of the ocean ecosystem, including plankton and benthic fauna (Blackford,  
157 1997), see Table 1. The model divides phytoplankton into four functional types based on size: pico-  
158 phytoplankton, nanophytoplankton, microphytoplankton and diatoms (Baretta et al., 1995). ERSEM  
159 uses variable stoichiometry for the simulated plankton groups (Baretta-Bekker et al., 1997; Geider

160 et al., 1997) and represents the biomass of each functional type in terms of chlorophyll, carbon, ni-  
 161 trogen, and phosphorus, with diatoms also being represented by silicon. ERSEM predators consist  
 162 of three types of zooplankton (mesozooplankton, microzooplankton, and heterotrophic nanoflagel-  
 163 lates), with organic material being decomposed by a single type of heterotrophic bacteria (Buten-  
 164 schön et al., 2016). The model represents three different sizes of detritus (small, medium and  
 165 large) and three types of dissolved organic matter (DOM: refractory; semi-labile; labile). The in-  
 166 organic component of ERSEM includes nutrients such as nitrate, phosphate, silicate, ammonium,  
 167 and carbon, as well as dissolved oxygen. The carbonate system is also included in the model (Ar-  
 168 tioli et al., 2012). ERSEM has been used for many applications including NWES and Mediter-  
 169 ranean Sea biogeochemistry reanalyses (Ciavatta et al., 2016, 2018, 2019), NWES operational fore-  
 170 cast (Skákala et al., 2018) (Skákala et al., 2018; McEwan et al., 2021), and NWES climate projections  
 171 (Wakelin et al., 2015, 2020; Galli et al., 2024; McEwan et al., 2021) (Wakelin et al., 2015, 2020; Galli et al., 2024)  
 172 .

Functional Group	Class/Type	Chemical Components
Phytoplankton	Diatoms	<i>Chl</i> , C, N, P, Si
Functional Types (PFT)	Microphytoplankton	<i>Chl</i> , C, N, P
	Nanophytoplankton	<i>Chl</i> , C, N, P
	Picophytoplankton	<i>Chl</i> , C, N, P
Zooplankton	Mesozooplankton	C
	Microzooplankton	C, N, P
	Heterotrophic Flagellates	C, N, P
Bacteria	-	C, N, P
Detritus	Small	C, N, P
	Medium	C, N, P, Si
	Large	C, N, P, Si
Dissolved Organic Matter (DOM)	Labile	C, N, P
	Semi-labile	C
	Refractory	C
Nutrient	Nitrate ( $NO_3^-$ )	N
	Phosphate ( $PO_4^{3-}$ )	P
	Ammonium ( $NH_4^+$ )	N
	Silicate ( $SiO_4^{4-}$ )	Si
Other	Temperature	-
	Oxygen $O_2$	-

Table 1: Reference table for ERSEM pelagic variables used in this study. Chemical components are represented by the following symbols: *Chl* is chlorophyll; C is carbon; N is nitrogen; P is phosphorus and Si is silicon. Note that we also use total chlorophyll (denoted as *c* in this paper), which is a diagnostic variable calculated as the sum of chlorophyll concentrations from all PFT classes.

173 The coupler known as the “Framework for Aquatic Biogeochemical Models” (FABM) (Bruggeman  
 174 and Bolding, 2014) allows for the smooth combination of hydrodynamic and biogeochemical models,  
 175 and is used to couple GOTM with ERSEM in this work. The coupling of GOTM to marine BGC  
 176 models using FABM has allowed for a wide range of applications that include modelling of phytoplank-  
 177 ton growth (Kerimoglu et al., 2021), [examining the](#) implications of sea-ice BGC for oceanic emissions  
 178 (Hayashida et al., 2017), [investigations on assessing](#) the highly intermittent spatial variability of phy-  
 179 toplankton on sub-grid scales (Mandal et al., 2016), and enhancing stoichiometry in existing BGC  
 180 models (Anugerahanti et al., 2021).

### 181 2.3 Model configuration and synthetic data setup

182 We configure the GOTM-FABM-ERSEM setup for two different locations in the English Channel (see  
 183 Fig. 1) and use synthetic observations of each. The first location, known as L4 (50.25°N, 4.217°W),  
 184 is a highly biologically productive site with seasonally stratified dynamics (Pingree and Griffiths,  
 185 1978), influenced significantly by the outflow of the nearby Tamar and Plym rivers. Nitrate acts as  
 186 the primary limiting nutrient for phytoplankton growth. It is monitored by the Western Channel  
 187 Observatory (WCO) (<https://www.westernchannelobservatory.org.uk/>) and SmartSound Plymouth  
 188 (<https://www.smartsoundplymouth.co.uk/>).

189 Besides the L4 site, we configure a setup for an additional location, that we shall refer to as the  
 190 Central Western English Channel (CWEC), at 49.40°N, 4.217°W. This point is less biologically pro-  
 191 ductive and it is much less influenced by riverine outflow than L4. These differences are evident when  
 192 looking at the distributions of biogeochemical signals in the models applied at these two locations (see  
 193 Fig. A.1). The differences make CWEC a reasonable alternative test site for assessing the application  
 194 of the ML model, and its suitability to generalise the results of this study under different marine BGC  
 195 conditions.

196 The physical and biogeochemical models for each location are forced with data appropriate for the  
 197 study area, using the following datasets: the General Bathymetric Chart of the Oceans 2023 (1/240°  
 198 resolution) for water depth; the ECMWF ERA5 dataset (0.25°/hourly resolution) for meteorology;  
 199 the TPXO9-atlas (1/30° resolution) for tides; and the World Ocean Atlas 2018 (0.25° resolution) for  
 200 temperature, salinity and nutrient fields ([nitrate, phosphate and silicate](#)) for biogeochemical relaxation  
 201 profiles. A nutrient relaxation timescale of 3 months towards the World Ocean Atlas data is required  
 202 to prevent significant trends forming that cause the 1D model to gradually accumulate nutrients.  
 203 This relaxation is significantly longer than the assimilation cycle of 7 days, and so has little impact  
 204 on forecast errors ~~—at the surface. However, the relaxation profiles could contribute to controlling the~~  
 205 ~~sub-mixed layer in our setup (which is not updated during assimilation), which could help to mitigate~~  
 206 ~~some long-term biases in these areas. This is a potential limitation for operational scale systems which~~  
 207 ~~do not have or use these relaxation profiles.~~

208 Ensemble runs, whether as free runs or for the EnKF (see Sect. 2.4.2) are configured and run using  
 209 the Ensemble and Assimilation Tool (EAT) in Python (Bruggeman et al., 2024). Each ensemble is  
 210 given a spin-up period of 10 years to settle the biogeochemistry appropriately and provide well-spread  
 211 initial conditions. Each ensemble member uses a signal of temporally correlated random noise to scale  
 212 the ECMWF ERA wind forcing at the location. The ~~scaling noise signal noise signal used to scale the~~  
 213 ~~wind forcing~~ has a correlation timescale of 7 days, a mean of 1 and a standard deviation of ~~0.30.5~~. The  
 214 resulting variation in wind strength across the ensemble members increases their spread over time,  
 215 and prevents ensemble collapse ([at least within the mixed layer](#)) induced by the previously mentioned  
 216 nutrient relaxation, or lack of representation of ~~other~~-error growth processes like horizontal advection  
 217 ~~—~~which are absent in a 1D set-up.

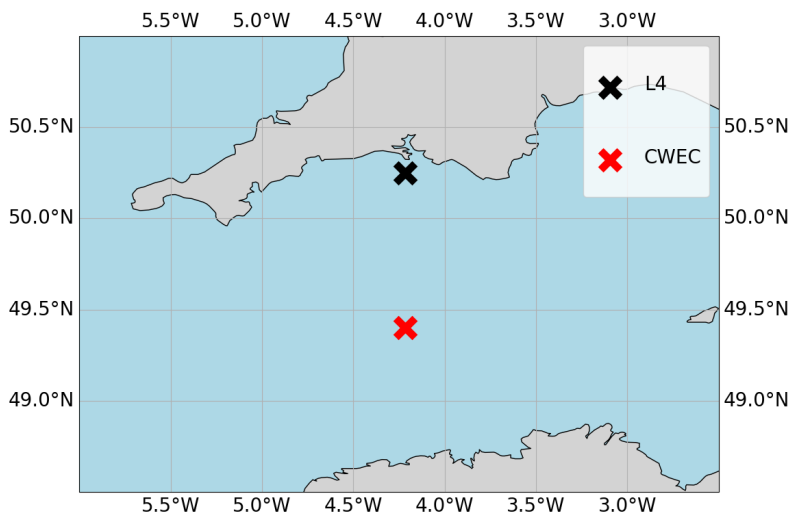


Figure 1: Map of the Western English Channel, marking the L4 model-training location with a black cross and the CWEC (Central Western English Channel) with a red cross, where we evaluated the model portability.

218 [For the purposes of training ML models, and generating climatological statistics, time periods for](#)  
 219 [the L4 location are partitioned as follows: training data \(2000-2014\), validation \(2015-2017\), offline](#)  
 220 [test \(2018-2020\), online test \(2022-2023\). Offline refers here to a setup in which the ML-OI analysis](#)  
 221 [is not then used as the initial condition for the next forecast, and so it does not impact successive](#)  
 222 [DA cycles. Conversely, online refers to a setup in which updates to the system can have dynamical](#)  
 223 [impact on later DA cycles as the model integrates forward in time. Climatological correlations and](#)

variances are calculated using a free-run ensemble over the training period. Because the forecasts extends seven days into the future, the number of available samples for any specific calendar day was limited, even though forecasts were issued throughout the full training period. To address this, the climatological statistics were computed as daily values, defined by averaging the statistics for a given day across all years in the training set. To increase the sample size and smooth out variability, a  $\pm 30$ -day window around each calendar day was applied. The CWEC location uses a run spanning 2000 - 2010 to generate climatological statistics, and the online test is performed for (2022-2023). No validation or offline test period is required for CWEC in this work, as we are primarily interested in a “naive transferability” of the model trained and evaluated at L4.

## 2.4 Data assimilation setups

We examine a total of five data assimilation (DA) setups in this work. These are conventional DA methods – namely a simple univariate scheme to reflect how DA is done currently in operational marine BGC systems, and an EnKF for comparison – to the new schemes that are hybridised with ML techniques. Before describing each scheme, we give the basic equations of conventional DA, introduce the state vector that the DA uses, the observation type that we will assimilate, and other adjustments that are done post assimilation.

The update equation that is central to DA is sometimes called the best linear unbiased estimator (BLUE, Asch et al., 2016; Carrassi et al., 2018) and is given by

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathcal{H}(\mathbf{x}^f)), \underbrace{\mathbf{K}(\mathbf{y} - \mathcal{H}(\mathbf{x}^b))}_{\text{analysis increment}} \quad (1)$$

where  $\Delta \mathbf{x}$  is the analysis increment

$$\Delta \mathbf{x} = \mathbf{K}(\mathbf{y} - \mathcal{H}(\mathbf{x}^f)),$$

and  $\mathbf{K}$  is the Kalman gain matrix

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^\top (\mathbf{H} \mathbf{P}^b \mathbf{H}^\top + \mathbf{R})^{-1}, \quad (2)$$

and where  $\mathbf{x}^a$  is the analysis (updated) state,  $\mathbf{x}^f$   $\mathbf{x}^b$  is the background state (which in this work is equivalent to a seven-day forecast state),  $\mathbf{y}$  are the observations,  $\mathcal{H}$  is the observation operator (with Jacobian  $\mathbf{H}$ ),  $\mathbf{P}^f$   $\mathbf{P}^b$  is the background error covariance matrix, and  $\mathbf{R}$  is the observation error covariance matrix. The matrix  $\mathbf{P}^f$   $\mathbf{P}^b$  is of special interest to this work. Ideally this matrix should be appropriately flow-dependent, but in practice it is often not, such as in many operational schemes. The purpose of this work is to introduce such flow-dependency to  $\mathbf{P}^f$   $\mathbf{P}^b$ , or to the analysis increments  $\Delta \mathbf{x}$ , with ML techniques.

In this work, the state of vector of the DA system,  $\mathbf{x}^b$  and  $\mathbf{x}^a$ , consists of the system, both  $\mathbf{x}^f$  and  $\mathbf{x}^a$ , comprises the surface values of most pelagic variables in ERSEM. We observe only the total chlorophyll,  $x_c$ , total chlorophyll and a set of chosen unobserved pelagic ERSEM variables (with chosen variable setups detailed in Sect. 3). While the state vector only considers surface values (OD), the entire DA process itself is 1D, using an idealised structure function to spread increments uniformly throughout the mixed layer. This approach assumes that surface conditions are representative of the mixed layer as a whole (which is broadly true in this model configuration), so that increments derived at the surface – However, total chlorophyll also provide a reasonable correction at depth within the mixed layer. We do not update the variables below the mixed layer, as they are decoupled or weakly coupled with the surface. Extending increments deeper would risk introducing spurious vertical structures, distorting stratification, or interfering with biogeochemical processes that are driven by different controls (e.g., remineralisation). By restricting updates to the mixed layer, the assimilation scheme ensures consistency with the available observations and avoids imposing unsupported corrections in the sub-mixed layer. Strategies for updating below the mixed layer may therefore require additional observation types (e.g., from floats or sea-gliders) or bias-correction approaches, rather than relying solely on surface observations combined with DA.

We assimilate only total chlorophyll,  $y$ , at the surface. Total chlorophyll in the model,  $x_{chl}$ , is a diagnostic variable constructed obtained by summing the chlorophyll concentrations of each phytoplankton function type (PFT) content of all phytoplankton functional types (PFTs, see Table 1) that are, which are themselves prognostic variables.

We will only ever have a single surface observation of total chlorophyll at each assimilation step. Our observation operator is therefore always linear ( $\mathcal{H}(\bullet) = \mathbf{H}\bullet$ ) and takes the form:

$$\mathbf{H} = [1, 0_1, \dots, 0_N],$$

273 In principle, one could keep only the PFT chlorophyll concentrations in the state and represent total  
 274 chlorophyll via the observation operator. However, this would require explicitly summing across  
 275 PFTs at every assimilation step, making the DA equations more cumbersome. Instead, we treat  
 276 surface total chlorophyll directly as part of the state vector. This simplifies the presentation of the  
 277 analysis equations, since the observation operator then reduces to a simple selection operator:

$$\mathbf{H} = [1, 0_1, \dots, 0_N], \quad (3)$$

278 where  $N$  is the number of chosen unobserved variables remaining in the state, and so the dimension  
 279 of the system is with the state ordered as  $(x_{chl}, x_1, \dots, x_N)$ . The system dimension is therefore  $N + 1$ ,  
 280 with the first element being corresponding to surface total chlorophyll. Note that this, and the  
 281 remaining elements corresponding to the unobserved variables to be updated.  $\mathbf{H}$  has the structure of  
 282 is a row vector as we have only a single observation for each, as there is only one observation per DA  
 283 update.

284 The PFTs themselves are excluded from the state, as they are and instead updated after the  
 285 analysis main analysis step in Eq. (1). The PFT chlorophyll updates are computed as a proportion  
 286 of the background ratio with total chlorophyll: surface total chlorophyll increment is redistributed  
 287 across PFTs in proportion to their background contribution to  $x_{chl}$ :

$$x_\chi^a = x_{\chi^-}^{fb} + \frac{x_\chi^f}{x_c^f} \frac{x_\chi^b}{x_{chl}^b} \cdot \left( x_{chl}^a - x_c^f - x_{chl}^b \right), \quad (4)$$

288 where  $\chi$  stands for each of the chlorophyll components of the PFTs, and  $c$  stands for the total  
 289 chlorophyll. Then, the additional the chlorophyll component of each PFT.

290 The associated chemical components of each PFT, which are carbon, nitrate, phosphate (and in  
 291 the case of diatoms, silicate as well) are updated according to the background ratios between the  
 292 chlorophyll of the given PFT and the chemical component as: (C, N, P, and for diatoms also Si) are  
 293 then updated in proportion to their background stoichiometric ratios with PFT chlorophyll:

$$x_\zeta^a = x_{\zeta^-}^{fb} + \frac{x_\zeta^f}{x_\chi^f} \frac{x_\zeta^b}{x_\chi^b} \cdot \left( x_\chi^a - x_{\chi^-}^{fb} \right), \quad (5)$$

294 where  $\zeta$  stands for each of represents the non-chlorophyll chemical components of a PFT. Which of the  
 295 remaining elements of the state vector at the surface are updated, and how, depends on the specific  
 296 DA scheme (see below).

297 All DA methods here will only directly update the surface layer components of each PFT. This  
 298 constrained redistribution scheme (Teruzzi et al., 2014; Skákala et al., 2018) ensures that phytoplankton  
 299 updates preserve forecast stoichiometric ratios and remain physiologically consistent, rather than  
 300 allowing the EnKF to update PFTs freely. While an unconstrained EnKF would eventually converge  
 301 towards similar balances, this explicit approach guarantees that assimilation respects acclimation  
 302 dynamics and maintains the community structure of the model. However, the rest of the mixed  
 303 layer is also updated, and uses this same surface analysis increment. Propagating surface analysis  
 304 increments in this manner makes the reasonable assumption that the behaviour at the model surface  
 305 is largely representative of behaviour through the mixed layer, and so the resulting analysis increments  
 306 at the surface are also approximately correct across these additional depth layers. Variables below the  
 307 mixed layer are not updated. Importantly, the same ratio-based balancing scheme can also be applied  
 308 outside an ensemble framework in single-model runs, where it provides a consistent way of updating  
 309 PFTs from a total chlorophyll correction.

310 The above observation operator, and updates to the PFTs are used in all DA schemes in this  
 311 paper. The specific DA schemes (conventional and ML-based) are now described. A summary of the  
 312 methods is given in Table 2.

### 313 2.4.1 Reference univariate DA scheme (RUS)

314 We call our baseline DA method the reference “univariate” DA scheme (RUS, Table 2, row 4). Its  
 315 purpose is to mimic existing DA systems used by several operational centres (Teruzzi et al., 2014;  
 316 Skákala et al., 2018), although our scheme is not variational. The background error covariances are  
 317 based on climatological information and so do not adapt to the state.

318 The RUS is based on an evaluation of Eq. (1), but only to directly update the total chlorophyll  
 319 variable. The simple structure of the observation operator in Eq. (3) means we can rewrite the

Run / scheme	Description / purpose	<i>chl</i> variance	<i>i</i> variance	<i>chl-i</i> correlation source	$\Delta x_{chl}$	$\Delta x_i$
<b>Preparation / training runs</b>						
1. Truth run	To synthesise observations and for analysis evaluation	n/a	n/a	n/a	n/a	n/a
2. Ensemble of free-runs	To determine climatological correlations and training for ML-OI	n/a	n/a	n/a	n/a	n/a
3. EnKF	Update all chosen surface variables / gold standard run / training for ML-EtE	ensemble-based	ensemble-based	ensemble-based	Eq. (1)	Eq. (1)
<b>Conventional assimilation runs</b>						
4. RUS	Reference univariate scheme (TC + stoichiometrical PFT update) / baseline for extensions	climatology	n/a	n/a	Eq. (6)	zero
<b>RUS extension assimilation runs (update to variable <i>i</i> with ML methods)</b>						
5. ML-OI	ML correlation hybrid	climatology	climatology	ML of free run	As RUS	Eq. (7)
6. ML-EtE	ML end-to-end EnKF emulation	climatology	n/a	n/a	As RUS	ML
<b>RUS extension assimilation runs (update to variable <i>i</i> with non-ML methods)</b>						
7. CliC	Climatological correlations	climatology	climatology	climatology	As RUS	Eq. (7)

Table 2: An overview of the different run-types and schemes used in this work. [Index \*chl\* refers to total chlorophyll, while index \*i\* refers to an unobserved variable \(e.g., nitrate\).](#) The truth run refers to a single-model run with no updates. We sample synthetic observations from this run and feed these into each DA scheme. The ensemble of free-runs means the model is left to run without assimilation. The EnKF uses an ensemble to model background error covariance in the DA update of all state variables (Sect. 2.4.2). The RUS is the ‘univariate’ scheme (Sect. 2.4.1), which is used as a benchmark for the performance of other schemes. It updates only the total chlorophyll state variable. The ML-OI [predicts-estimates](#) background correlations of variables beyond the total chlorophyll with an ANN (Sect. 3.1). The ML-EtE [predicts-estimates](#) the analysis increments of variables beyond the total chlorophyll produced by an EnKF using an ANN (Sect. 3.2). The CliC is similar to ML-OI but uses purely climatological background statistical estimates of the correlations to update the state of unobserved variables (Sect. 3.3).

320 update Eq. (1) to show how the total chlorophyll (index [\*echl\*](#)) is updated from the total chlorophyll  
321 observation:

$$x^a_{\underline{echl}} = x^{\text{fb}}_{\underline{echl}} + \frac{P^f_{c,c}}{P^f_{c,c} + R} \frac{P^b_{chl,chl}}{P^b_{chl,chl} + R} \cdot (y - x^{\text{fb}}_{\underline{echl}}), \quad (6)$$

322 where  $\frac{P^f_{c,c}}{P^f_{c,c} + R} \frac{P^b_{chl,chl}}{P^b_{chl,chl} + R}$  is the background error variance of total chlorophyll [and \*R\* is the observation error](#)  
323 [variance](#). Climatological variances from a long training EnKF run (Sect. 2.4.2) are used to estimate  
324  $\frac{P^f_{c,c}}{P^f_{c,c} + R} \frac{P^b_{chl,chl}}{P^b_{chl,chl} + R}$ . Details on the training runs can be found in Sect. 3.1 and 3.2. Updates to the surface  
325 PFT chlorophyll, to the associated chemical components, and throughout the mixed layer are made  
326 separately as described [at the start of previously in](#) Sect. 2.4.

327 Note that we call this scheme “univariate” as only a single variable (total chlorophyll) is updated  
328 according to the background and observational errors as described in Eq. (6). All further DA schemes  
329 described in this work (apart from the EnKF below, which uses ensemble-derived covariances) start

330 with an update of the total chlorophyll using Eq. (6), and will attempt to update additional pelagic  
 331 variables using the new ML-based approaches.

### 332 2.4.2 The EnKF-based scheme

333 The stochastic EnKF scheme, see e.g., Evensen (2003), approximates the update Eqs. (1) and (2) with  
 334 an ensemble to estimate the flow-dependent background error covariance matrix  $\mathbf{P}^f \mathbf{P}^b$ , Table 2, row  
 335 3. For each ensemble member there is a different update and a different perturbed observation (the  
 336 perturbations are sampled from the normal distribution  $\mathcal{N}(0, R)$ ). The EnKF updates all elements of  
 337 the surface state described previously using the ensemble version of Eq. (1), but still performs the sto-  
 338 ichiometric balancing scheme and duplication of the analysis increments from the surface throughout  
 339 the mixed layer, as described in Sect. 2.4. This is done so that there is a one-to-one correspondence  
 340 between the strategy used to generate the training data, and the strategy applied in the single-model  
 341 schemes.

342 ~~The implementation of an EnKF to this problem is relatively expensive, but provides a “gold~~  
 343 ~~standard” comparison DA method, plus it provides valuable training data for the ML model, as~~  
 344 ~~specified in Sect. 3.2. The accuracy of the EnKF will depend on the number of ensemble members.~~  
 345 ~~This is discussed in Sect. 4.2.~~

## 346 3 Hybrid machine learning data assimilation for marine bio- 347 geochemistry

348 In this section, we describe how we hybridise the DA, described above, with ML to provide flow-  
 349 dependent estimates of the statistics/increments that are better than the climatological values. In  
 350 particular, we take two approaches that differently replace parts of, or fully, the update equation.  
 351 We now show the mathematical framework that the ML schemes will emulate, which is derived from  
 352 Eqs. (1)-(3).

### 353 3.1 ~~Mathematical framework for the ML-based DA schemes~~

354 The ML-based DA schemes are summarised in Table 2, rows 5 and 6. They both build upon RUS,  
 355 extending it to become multivariate. The total chlorophyll analysis is computed using the RUS update  
 356 Eq. (6), while the remaining variables (potentially  $1 \leq i \leq N$ ) have updates according to

$$x_i^a = x_i^{fb} + \underbrace{\frac{P_{i,chl}^b}{P_{chl,chl}^b + R} \cdot (y - x_{chl}^b)}_{\text{analysis increment}}, \quad (7)$$

357 where  $\frac{P_{i,c}^f}{P_{i,chl}^b}$  is the background error covariance between variable  $i$  and total chlorophyll defined  
 358 as

$$P_{i,chl}^{f,b} = \text{COR} \rho_{i,chl} \cdot \sigma_i \cdot \sigma_{chl}, \quad (8)$$

359 where  $\text{COR}$  is their forecast error correlation, and  $\sigma_i$  and  $\sigma_{chl}$  are their  
 360 respective background error standard deviations. In Eq. (7) the analysis increment of the update is  
 361 labelled.

362 An important aspect of any DA scheme is its ability to adapt with the flow. A conventional  
 363 way to introduce flow-dependency is via Monte Carlo-like methods such the EnKF, which comes with  
 364 substantial computational cost. The two proposed ML-DA schemes below are designed with the above  
 365 in mind and provide flow-dependency cost-effectively without the need for an ensemble (apart from  
 366 at the training stage, as shall be clarified). The two ML-DA schemes are described in Sections 3.1  
 367 and 3.2. ~~The specific details of their architecture are given in-~~

368 Regardless of the specific ML-DA scheme, each ML model is a fully connected ANN optimized using  
 369 AutoKeras (Jin et al., 2019) over 100 trial configurations. AutoKeras uses Bayesian optimisation in  
 370 a network search algorithm to determine optimal hyperparameters such as layer depth, layer width,  
 371 dropout rate, learning rate, and optimiser selection. The input features of each model are standardised  
 372 to have unit variance and a mean of zero, using data from the L4 training period during training.

373 Each ML approach is tested in the following scenarios: (1) a set-up where the only unobserved  
 374 variable that is updated is nitrate, (2a) a set-up where we update the full set of pelagic variables,  
 375 and (2b) a set-up where we update a partial set of the pelagic variables, eliminating poorly estimated

variables based on the results of (2a). The progression from (1) to (2a) allows us to move from a controlled test of a single key limiting variable to a comprehensive update of the full system, while (2b) represents a refinement step that is only possible after evaluating the performance of (2a). In this way, the experimental design not only tests the limits of updating all variables, but also demonstrates how excluding problematic variables can improve robustness without discarding the broader benefits of multivariate updates. In each setup, the number of outputs to be estimated by the ML model corresponds to the number of unobserved variables. In the case of ML-OI (see Sect. 3.1), the standard deviations must also be estimated from climatology.

### 3.1 Hybrid machine-learning optimal interpolation (ML-OI)

This approach first updates the observed total chlorophyll and associated PFTs in an identical manner to the RUS described in Sect. 2.4.1. The first ML method uses the sealed background state (see below) to predict the state-dependent correlations  $\rho_{i,chl}$  between observed and unobserved quantities in Eq. (8) as a function of the background state (even though the EnKF update of state variables from other state variables is linear, the relationship between state variables and correlations is likely non-linear). Together with climatologically estimated values of  $\sigma_i$  and  $\sigma_c \sigma_{chl}$  (estimated using a free-run ensemble over the training period as described in Sec. 2.3), the correlations are substituted into Eqs. (8) and then (7) to provide updates to the unobserved variables in the system. We call this approach ML-OI (“optimal interpolation”, Table 2, row 5). For each variable input into the ML-OI model to predict  $COR_{i,c}$  is sealed according to estimate the correlation, the background state is additionally divided by its climatological maximum for the given location (e.g., L4 in Fig. 1), assuming that the correlative relationship between variables will be similar, even if the scale of seasonal variability is different. Standard deviations are each a statistic of a single variable, which we assume is easier to capture climatologically than a correlation is. The at the corresponding location (before the regular standardisation procedure is applied to the data). This normalization accounts for differences in the amplitude of seasonal variability while making a bold assumption that the underlying correlative relationships between variables remain consistent across locations. Since correlations are dimensionless, this scaling does not affect their interpretation, and the subsequent analysis increments (which are constructed by combining the estimated correlations with the location-specific climatological variances) remain physically consistent. Since variances are single-variable statistics that can be estimated more reliably than correlations from long climatological records, we assume they are sufficiently robust to provide a stable basis for use in the Kalman gain. In contrast, correlations describe joint variability and therefore require the more sophisticated, data-driven approach described above, and cannot be approximated in the same way.

As with every approach used in this work, the resulting surface increments of the unobserved variables are then propagated to the other levels in the mixed layer, as described previously.

While the field of hybrid ML-DA is growing rapidly, there exists relatively few works in which ML-predicted background error covariances are so closely coupled with existing DA systems. However, a few particularly relevant examples stand out such as Ouala et al. (2018), in which a Kalman-like analysis update is applied to satellite-derived sea surface temperature fields using artificial neural network (ANN)-predicted background error covariances. Additional examples of this can be seen in Sacco et al. (2022), which aim to learn different sources of uncertainty using ANNs on both toy models and sea level pressure forecasts. Further work (Sacco et al., 2024) uses an EnKF to generate flow dependent background error covariances, and then learns them using a convolutional neural network.

In order to generate training data for this approach, we run a 100-member ensemble of free-runs, configured according to Sect. 2.3 (Table 2, row 2). We generate training samples at seven day intervals across these free-runs, covering the period from 2000-2014. The features are the surface states of individual ensemble members at a given time, across all pelagic model variables. For the first application of ML-OI in Sect. 4.2, the targets are time dependent/ensemble-derived correlations between total chlorophyll and nitrate. In the later application in Sect. 4.3 onwards this is extended from just nitrate to a wider set of variables.

While the field of hybrid ML-DA is growing rapidly, there exists relatively few works in which ML-estimated background error covariances are so closely coupled with existing DA systems. However, a few particularly relevant examples stand out such as Ouala et al. (2018), in which a Kalman-like analysis update is applied to satellite-derived sea surface temperature fields using artificial neural network (ANN)-estimated background error covariances. Additional examples of this can be seen in Sacco et al. (2022), which aim to learn different sources of uncertainty using ANNs on both toy models and sea level pressure forecasts. Further work (Sacco et al., 2024) uses an EnKF to generate

434 flow dependent background error covariances, and then learns them using a convolutional neural  
435 network.

## 436 3.2 End-to-end machine learning of EnKF updates (ML-EtE)

437 This approach again first updates the observable-observed total chlorophyll and associated PFTs in an  
438 identical manner to the RUS described in Sect. 2.4.1. Nevertheless, as opposed to ML-OI, ML is used  
439 here to predict-estimate the analysis increments for unobserved variables, given the analysis increment  
440 of the observed variable (total chlorophyll) and the complete background state. This obviously requires  
441 running a DA system to learn from. This is achieved here using the updates produced by an EnKF  
442 training run (see below). We call this approach ML-EtE (“end-to-end”, Table 2, row 6) emulation of  
443 an existing DA system.

444 In ML-EtE, ~~we assume that the essential properties of each statistical object that creates an~~  
445 ~~analysis increment, such as covariances and observation uncertainty, can be more effectively captured~~  
446 ~~by directly predicting the analysis increment rather than predicting every component individually and~~  
447 ~~allowing errors to compound across multiple independent predictions that~~ the analysis increment is  
448 predicted directly. By contrast, ML-OI predicts correlations, which are then combined into a single  
449 value. The resulting surface increments of the unobserved variables are then propagated to the other  
450 levels in the mixed layer as described previously.

451 This approach follows other work in a similar vein, such as work by Bonavita and Laloyaux (2020)  
452 , which uses ANNs to emulate the main features of an operational weak-constraint 4D-Var scheme.  
453 ~~Let us reiterate however that ML-EtE replaces, in an end-to-end fashion, only the analysis step of~~  
454 ~~the EnKF, and not the full forecast-analysis cycle (including therefore the dynamical model).~~ As  
455 mentioned in the introduction, a similar scope characterises the work by Boequet et al. (2024) with  
456 climatological standard deviations to form a Kalman gain, and only then applied to the innovation  
457 to yield the increment. The ML-OI scheme introduces potential sources of error - both from the  
458 uncertainty of data-driven correlation estimates and from the reliance on climatological statistics.  
459 Directly estimating the analysis increment (a vector) is also naturally more scalable for high-dimensional  
460 applications (e.g., operational 3D systems), where manipulating the full error covariance matrix - or  
461 even reduced or reformulated versions - becomes computationally costly.

462 To generate the training data for this approach, we first generate a nature run for the training  
463 period (Table 2, row 1), to generate synthetic surface observations of total chlorophyll concentration  
464 at weekly intervals. The observation uncertainty is equal to 10% of the observed value. These are then  
465 assimilated into the EnKF run over the same period. The features of each training sample consist of an  
466 individual ensemble member’s background state and its corresponding total chlorophyll increment from  
467 the EnKF run. The targets are the corresponding analysis increments for the unobserved variables –

468 ~~As mentioned, emulating analysis increments of existing DA systems is shown in existing at the~~  
469 ~~surface.~~ As described previously, the resulting surface increments of the unobserved variables are then  
470 propagated to the other levels within the mixed layer.

471 ~~This approach follows other non-marine BGC works, in particular in relation to estimating and~~  
472 ~~correcting model error (Bonavita and Laloyaux, 2020; Brajard et al., 2020; Gregory et al., 2024).~~ In  
473 all cases, it relies on having work in a similar direction, such as Bonavita and Laloyaux (2020), who  
474 used ANNs to emulate the main features of an operational weak-constraint 4D-Var scheme, while  
475 Bocquet et al. (2024) pursued a similar end-to-end replacement of the analysis step. Likewise, several  
476 studies have demonstrated the emulation of analysis increments to estimate and correct model error  
477 (Brajard et al., 2020; Gregory et al., 2024). A common feature of these works is their reliance on an  
478 existing DA system in place, or, as in the above mentioned works or, more generally, on a robust  
479 reanalysis. The availability of a reanalysis is a clear obstacle need for such a reanalysis represents  
480 a key limitation of this approach . We shall discuss this further in our — one we revisit in the  
481 conclusion. Here we are primarily interested in studying its feasibility and, however, our primary  
482 goal is to examine the feasibility of ML-EtE and its ability to learn successfully the EnKF updates  
483 effectively. Using this approach, the increments still ultimately come from the “analysis-background”  
484 of an EnKF, which provides a linear update to the system (even if the relationship between the state  
485 at the analysis increments is non-linear). Going beyond this limitation would require either training  
486 on increments relative to the true state (e.g., truth-background) or employing a non-linear DA system,  
487 such as a particle filter, to capture more complex, non-linear corrections.

## 488 3.3 Machine learning architecture

489 Each ML model is a fully connected ANN optimized using AutoKeras (Jin et al., 2019) over 100 trial  
490 configurations. AutoKeras uses Bayesian optimisation in a network search algorithm to determine  
491 optimal hyperparameters such as layer depth, layer width, dropout rate, learning rate, and optimiser  
492 selection. The input features of each model are standardised to have unit variance and a mean of  
493 zero.

494 Each ML approach is tested in the following scenarios: (1) a set-up where we choose to update  
495 only nitrate, (2a) a set-up where we update the full set of pelagic variables, and (2b) a set-up where  
496 we update a partial set of the pelagic variables, eliminating poorly estimated variables based on the  
497 results of (2a).

### 498 3.3 Purely climatological updates

499 A further non-ML-based scheme is used to update the ~~extended range of variables~~ nitrate to mirror  
500 ML-OI, but using only climatological correlations derived from the EnKF run (CliC, Table 2, row 7)  
501 over the training period. This serves as another comparison point, a benchmark, to check whether  
502 the additional complexity of an ML model is needed.

### 503 3.4 Skill metric and machine learning model evaluation

#### 504 3.4.1 Skill metric

505 For a system that runs for  $\tau$  cycles (where a cycle represents a complete 7-day forecast and analysis),  
506 we represent the trajectory for a member  $i$  of ensemble  $X$  at cycle  $t$  as  $X_t^i$ . The truth is denoted as  $T_t$ .  
507 The expected RMSE (root mean square error) over  $M$  ensemble members (or a set of  $M$  single-model  
508 runs), is calculated as:

$$RMSE = \frac{1}{M} \sum_{i=1}^M \sqrt{\frac{1}{\tau} \sum_{t=1}^{\tau} (X_t^i - T_t)^2}. \quad (9)$$

509 This is a sensible metric to use when calculating the expected error across a set of independent single-  
510 model runs, such as in the RUS, CliC, ML-EtE and ML-OI approaches. It is also convenient for  
511 calculating the expected error of the ensemble members in the EnKF runs.

#### 512 3.4.2 SHAP analysis

513 Shapley values are a well known and widely used metric for understanding the importance and con-  
514 tributions of individual input features in ML models (Lundberg and Lee, 2017). A Shapley value  
515 represents the average marginal contribution of a feature across all possible subsets of features, ensur-  
516 ing a fair allocation of importance. In this work, we use Kernel SHAP (SHapley Additive exPlanations)  
517 as a model-agnostic approach to estimate mean Shapley values across a dataset. Kernel SHAP ap-  
518 proximates Shapley values by training a weighted linear model on perturbations of the input data. By  
519 calculating the mean absolute Shapley values, we measure the magnitude of influence for individual  
520 features to the model's ~~predictions~~ estimations.

521 By understanding the importance of each input feature, we gain insight into the correlative links  
522 of dynamical behaviour in the system. This can help to identify how-and-when a model will translate  
523 well to new conditions. For example, if the primary predictive feature of an ML model is similar in two  
524 separate locations, one trained and one unseen, then we may expect the ML model to perform reason-  
525 ably well in the new scenario, even if the other non-predictive features exhibit an entirely different  
526 distribution. We emphasise that we cannot infer causality from this analysis alone but understanding  
527 the data-driven feature importance and feature contribution for an ML model, combined with expert  
528 understanding of the system dynamics, can help to unveil connections and insights into the complex  
529 processes of the marine BGC model.

530 It also also worth noting that these metrics can also be used for feature selection with the idea  
531 that if a feature contributes little-to-nothing to the predictions, it can probably be eliminated from  
532 the feature set. This then requires the expensive processes of iteratively re-training and re-testing  
533 the neural networks and so is not an avenue that we explore in this work. SHAP is somewhat  
534 limited in the presence of highly correlated features because Shapley values assume independent feature  
535 contributions. This can lead to arbitrary or shared attributions when features provide redundant  
536 information, making it difficult to disentangle their true individual impacts. However, the correlation  
537 structures of the ~~MGBC~~ marine BGC have been studied previously (Higgs et al., 2024), and so can  
538 be more effectively accounted for during analysis.

## 4 Results and discussion

### 4.1 System dynamics

As discussed in Sect. 2.3, the L4 location is a highly biologically productive site with seasonally stratified dynamics. Nitrogen is a key component of organic matter and is generally the limiting nutrient to primary production by phytoplankton in coastal marine ecosystems (Council et al., 2000), which includes the L4 location (Smyth et al., 2010). This leads to a strong, potentially exploitable dynamical link between phytoplankton and nitrate that varies with a clear seasonal cycle. This cycle Figure 2 demonstrates this seasonality, and how it can be broken down into three distinct regimes across any given year:

- The *light-limited regime* typically describes a fully mixed water column, which approximately spans the period from October to the start of the next spring bloom. Here, there is little-to-no phytoplankton growth due to the reduced light-levels, meaning nutrients are mixed throughout the water column without being used by the phytoplankton. During this period, phytoplankton concentrations are very low and mostly decoupled from nutrient dynamics<sup>2</sup>.
- The *bloom regime* can occur throughout spring (from March until May), and is the period when phytoplankton reaches its yearly maximum. During this time, light levels no longer limit phytoplankton growth and there is a high availability of nutrients that have accumulated in the water column during the “light-limited” period. This results in a rapid increase of phytoplankton concentration, and an exhaustion of nutrients.
- The *nutrient-limited regime* refers to the period roughly spanning from early summer until late September where nutrients, and more specifically nitrate, have been exhausted by the phytoplankton during bloom and so concentrations are generally very low. During this time, phytoplankton relies on processes such as storms to mix nutrients into the upper water column. Consequently, phytoplankton growth is sporadic and less intense than during the spring bloom.
- ~~The light-limited regime typically describes a fully mixed water column, which approximately spans the period from October to the start of the next spring bloom. Here, there is little-to-no phytoplankton growth due to the reduced light-levels, meaning nutrients are mixed throughout the water column without being used by the phytoplankton. During this period, phytoplankton concentrations are very low and mostly decoupled from nutrient dynamics.~~

---

<sup>2</sup>The model is constrained to physically non-negative values, so any negative values that arise during the data assimilation (DA) step, though extremely rare, are clipped to zero. This occurs infrequently and only in very localized instances. As such, this treatment has a negligible impact on the overall state and statistical distributions. Additionally, due to strong surface forcing, the model tends to quickly redistribute any localized anomalies, minimizing the persistence or propagation of these clipped values. Therefore, we are confident that this approach does not significantly influence the model performance or results.

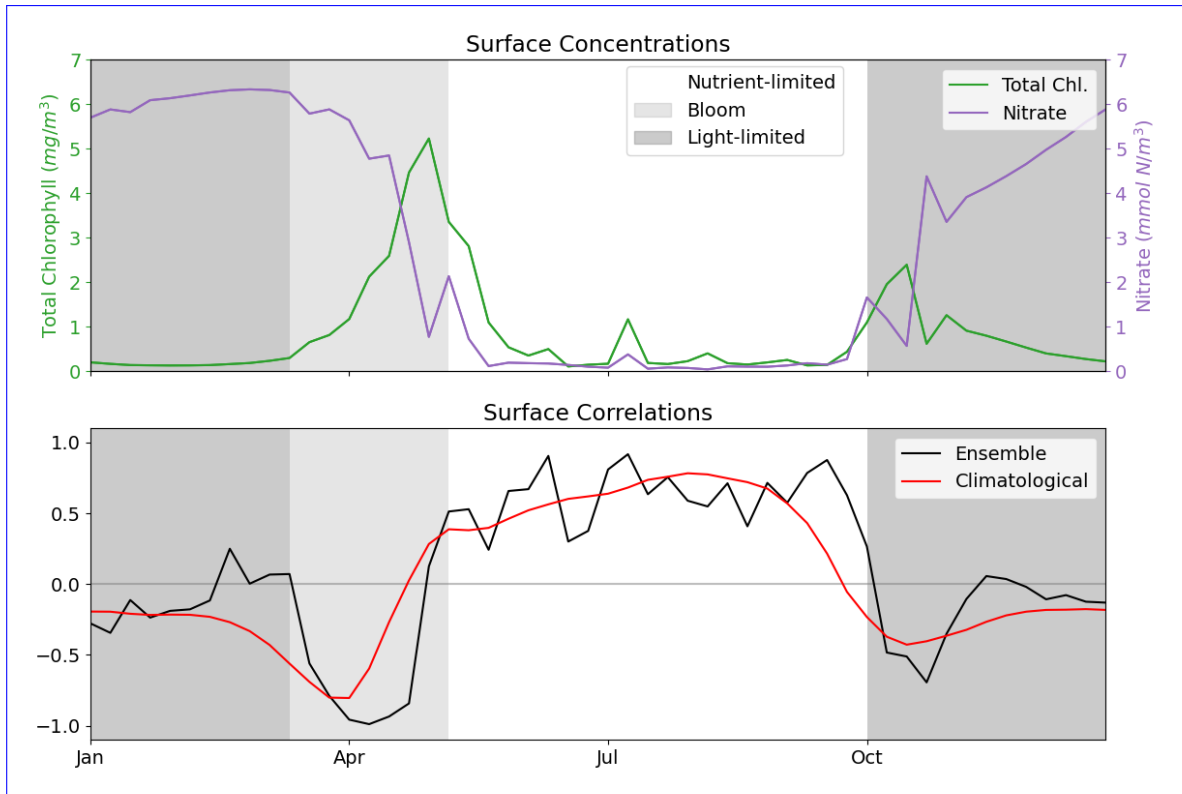


Figure 2: The top panel shows a time series of for the surface concentrations of total chlorophyll (black green) and nitrate (green purple) for an arbitrary year during 2023 at the L4 location. The bottom panel shows the climatological correlation presents correlations between total chlorophyll and nitrate, derived from a 96-member ensemble (black) and from a daily varying climatology (red; calculated across over the 2000-2014 2000-2014 training period). Shading indicates the dominant seasonal system regimes: “light-limited” (white dark grey), “bloom” (light grey) and “nutrient-limited” (dark grey white).

## 4.2 Prediction Estimation and update to a single pelagic variable

In this section, we explore the performance of ML-OI and ML-EtE in updating only nitrate as an unobserved variable. Recall however that the observed total chlorophyll and associated PFTs are updated according to the RUS scheme in Sect. 2.4.1. We choose nitrate for these initial experiments because it is a limiting nutrient at the L4 location (see Fig. A.2 in the Appendix) (see Fig. A.2 in the Appendix and Smyth et al., 2010), and therefore has a clear, explainable relationship with total chlorophyll as discussed in Sect. 4.1 (see also Fig. 2). Since nitrate is the key driver limiting primary production among nutrients, addressing it through DA can could have a significant knock-on effect on the whole model state. Thus (through dynamical evolution from the corrected state). Moreover, this also provides us a more understandable proof-of-concept with reduced complexity to analyse initially, before we later extend the updates to more than 30 additional pelagic variables (see Table 1) in a higher complexity scenario. However, as will become clear in Sec. 4.3, this strategy for assigning importance or priority to variables in the DA scheme does not necessarily correspond to the dynamical importance of a variable in the model, highlighting the need for specific results.

Figure 3 shows the correlation between total chlorophyll and nitrate as a function of time in the period 2018-2020 for the “offline” ML-OI experiment. Offline refers here to a setup in which the ML-OI analysis is not then used as initial condition for the next forecast, and so it does not impact successive DA cycles. The performance of ML-OI is compared to the “true correlation” computed over an ensemble of 100 members and to the correlation estimated using daily climatology.

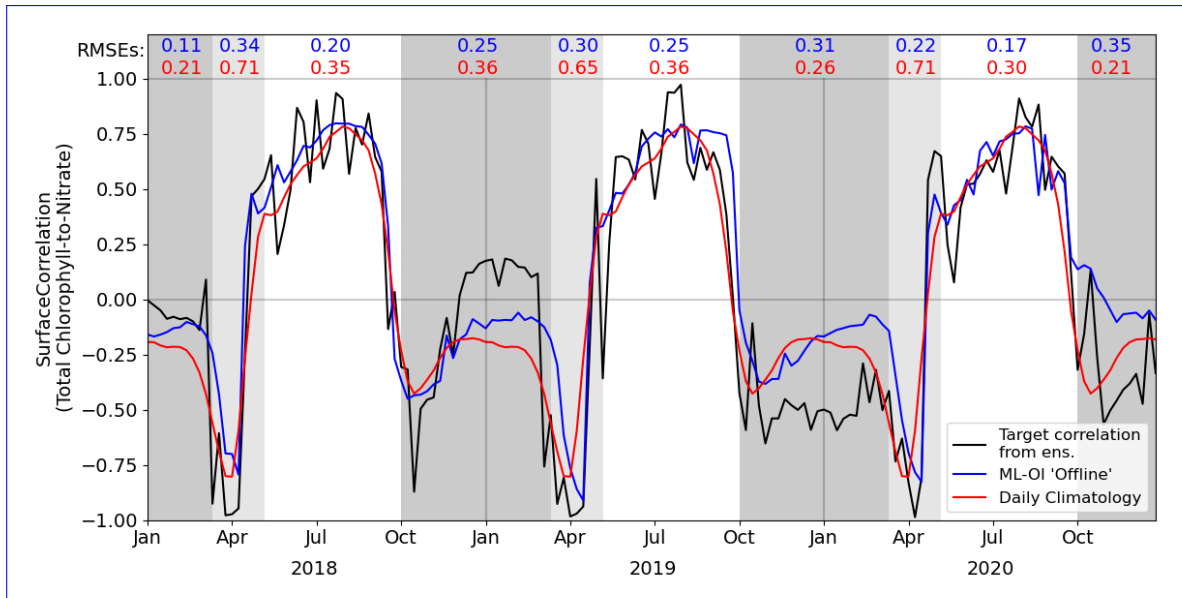


Figure 3: Predictions for Estimates of correlation between total chlorophyll and nitrate, at weekly intervals across the 3-year offline test period for the ML-OI approach. The “true” target correlation (black) is calculated from the 100-member free-run ensemble (Table 2, row 2). The predictions by Correlation estimates are shown for ML-OI (blue) are shown, with and the RMS difference to the true correlation of 0.255. A daily climatology of correlations has also been (red; calculated from over the 2000–2014 training data (red period). The root mean square errors between the estimated and target correlations are given at the top of the figure, calculated separately over each regime window and for both estimation methods (with an RMS difference of 0.731 corresponding colour). The seasonal regimes of Figure 2 are repeated.

587 The input features (i.e. model background states) and target correlations for  
 588 The ML-OI are taken from the 100-member free-run. As it will become clear below, this is a  
 589 sufficiently large ensemble to provide accurate correlation estimates. During this test period, the ML  
 590 model ( $RMSE = 0.255$ ) vastly outperforms the climatological statistics estimates ( $RMSE = 0.731$ ),  
 591 which is an approximate 60% improvement. model shows clear improvements over climatological  
 592 estimates of correlation across most of the annual cycle. It is important to note, however, that the  
 593 RMSE here only reflects the capability of a method to estimate the ensemble-derived chlorophyll-nitrate  
 594 correlations. Such improvements do not necessarily translate directly to performance in an online,  
 595 cycled data assimilation system (shown later) where past estimates can influence future states through  
 596 the model’s dynamical evolution.

597 The Most clearly, ML-OI model makes clear improvements over the climatological estimates at  
 598 several points in the yearly cycle. Firstly, it better estimates is better than climatology for estimating  
 599 the highly distinctive correlative pattern between total chlorophyll and nitrate during the bloom  
 600 regime, showing a moderate to significant RMSE reduction in every bloom period. This pattern  
 601 consists of a sharp drop to a strongly negative correlation, before an almost instantaneous increase to  
 602 a strong positive correlation. These correlation patterns can be simply explained. During the bloom,  
 603 phytoplankton growth exhausts nutrients, leading to negative correlations between chlorophyll and  
 604 nutrients, whilst the end of the bloom, and the following period, phytoplankton growth is nutrient  
 605 limited, leading to positive correlation. The precise timing of the bloom (and hence this correlation  
 606 pattern) has notable inter-annual variability – varying within a period of approximately 5 weeks each  
 607 year, in this model. The climatological correlations estimate this pattern poorly as they are smoothed  
 608 over this period of inter-annual variability, but the ML-OI model, which predicts estimates correlations  
 609 from the state of the marine BGC model, captures the pattern much more accurately. Also, during-

610 During the nutrient-limited regime, we see a generally strong positive correlation between total  
 611 chlorophyll and nitrate, which has some local variability primarily driven by changes in wind strength,  
 612 such as a weather front passing over the location and mixing nutrients into the surface. The ML scheme  
 613 captures ML-OI scheme clearly reduces the RMSE during this period, perhaps capturing some of the  
 614 local variability in this “true” signal and so responds more accurately to these changes in state. We  
 615 Though, it is clear that the correlations of the ensemble still vary more strongly than either other  
 616 method.

617 Finally, we also see that during the light-limited regime, the system can exist in either a “weakly  
 618 positive or no correlation” state, or a “moderately negative correlation” state. However, both the  
 619 climatological and ML estimate fail to capture these possible states. During this time, total chlorophyll  
 620 and nitrate are generally decoupled, and ~~the~~ there is there is no clear link between the state of  
 621 the system and the correlations estimated. Furthermore, as the ensemble concentrations in total  
 622 chlorophyll are very near zero, ~~meaning the data assimilation has~~ the DA updates are also small and  
 623 have very little impact ~~at this time of year.~~ The latter point means that better correlation predictions  
 624 This means that any improvement or degradation in correlation estimates at this time of year are  
 625 less likely to result in any great improvement to the system, as there is weak relationship between  
 626 chlorophyll and nitrate DA increments. However, updates at the start of this period ~~could~~ can be  
 627 important, as the resulting store of nitrate in the upper water column could have dynamical impact  
 628 in later DA cycles when light is no longer limiting and the next bloom period starts.

629 After demonstrating the capability of the ML model to ~~predict~~ estimate the chlorophyll-nitrate  
 630 relationship in an “offline” setting in Fig. 3, in Fig. 4 we compare the performance of a standard  
 631 EnKF at different ensemble sizes with the schemes previously summarised in Sect. 2.4 and 3. This  
 632 is done in an “online” setting, so that any update to the system can have dynamical impact on later  
 633 DA cycles as the model integrates forward in time.

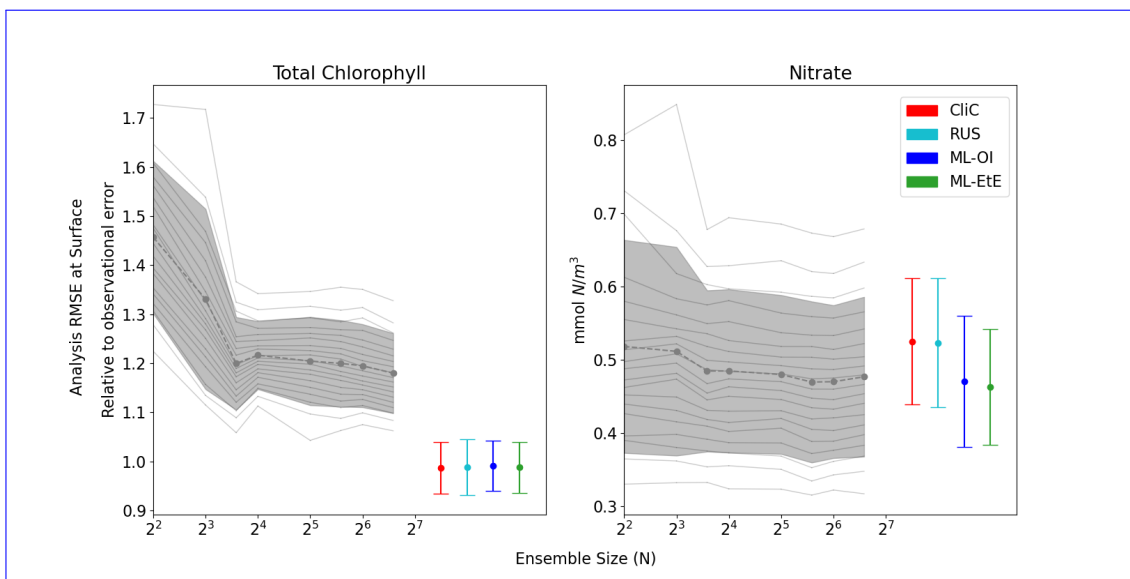


Figure 4: The relationship between analysis RMSE (Eq. (9)) and ensemble size for EnKFs with different ensemble sizes, as well as the performance of the different single-model run schemes. The left panel shows the RMSE of the observed variable, total chlorophyll, normalised relative to the observational error. The right panel shows the RMSE of the unobserved variable, nitrate. The black dashed line represents the mean expected ensemble member error, from an aggregated pool of ensemble members taken from 20 repeat experiments of an EnKF at increasing ensemble sizes, with the shaded grey area indicating  $\pm 1$  standard deviation of ensemble member error. The mean error and  $\pm 1$  standard deviation of 64 independent single-model runs are also given for each of the methods summarised in Sects. 2.4 and 3. An extended version of the plot, showing a wider range of unobserved, not updated variables is given in Fig.B.1.

634 Figure 4 displays the performance of the EnKF, the RUS scheme, the climatological statistics  
 635 scheme (CliC), and the ML schemes. Each panel shows the mean expected error of ensemble members  
 636 for ensemble sizes ranging from 4 to 96. In the left panel for total chlorophyll, the relative RMSE  
 637 is calculated as a percentage-ratio of the observation error. The EnKF achieves a near-optimal  
 638 performance at an ensemble size  $> 16$ , after the mean expected error of ensemble members reaches a  
 639 plateau with increasing size. The relative analysis error of total chlorophyll is normalised according  
 640 to observational error. This exceeds a value of 1 as we are measuring expected error of ensemble  
 641 members, not error to the ensemble mean, as described in Sect. 3.4.1. Since the EnKF generates  
 642 an ensemble of observations (with noise based on the uncertainty), an additional source of error is  
 643 introduced relative to the single-model runs which are not stochastic. This means that when we  
 644 calculate the error of each ensemble member, rather than the error of the ensemble mean, we get this  
 645 difference in error. We also see, in the right panel, that the error decreases with ensemble size for

646 the unobserved nitrate, indicating that the system converges towards more correct nitrate updates at  
647 larger ensemble sizes.

648 As expected, the analysis error in the observed total chlorophyll is generally comparable across  
649 each scheme because they all use the same method, the RUS scheme of Sect. 2.4.1, to update the  
650 observed total chlorophyll. However, there are more noticeable differences in the schemes that extend  
651 the updates to nitrate as well. In this, we can clearly see that both the RUS scheme (no update to  
652 nitrate) and the CliC scheme (update of nitrate using climatological covariances in Eq. (8)) perform  
653 similarly poorly in improving analysis error of nitrate – meaning that the information provided by  
654 the observation has not propagated well to the unobserved variable. In contrast to this, both ML  
655 approaches result in a significant improvement in performance, reducing analysis error by between 8 –  
656 12%. This means that the information from observations can effectively propagate to the unobserved  
657 variables in a single-model run, without the need for an expensive ensemble to model the statistics  
658 at run time. Also, this indicates that the improvement in correlation ~~prediction-estimation~~ shown  
659 in the offline experiment of Fig. 3 translates ~~into the entirety~~ (at least on average) into the online  
660 testing period, where the updates of a given DA cycle ~~will~~ feed into subsequent cycles. The standard  
661 deviation of ML-EtE is also lower than the spread for ML-OI, which is a good sign of lower sensitivity.

662

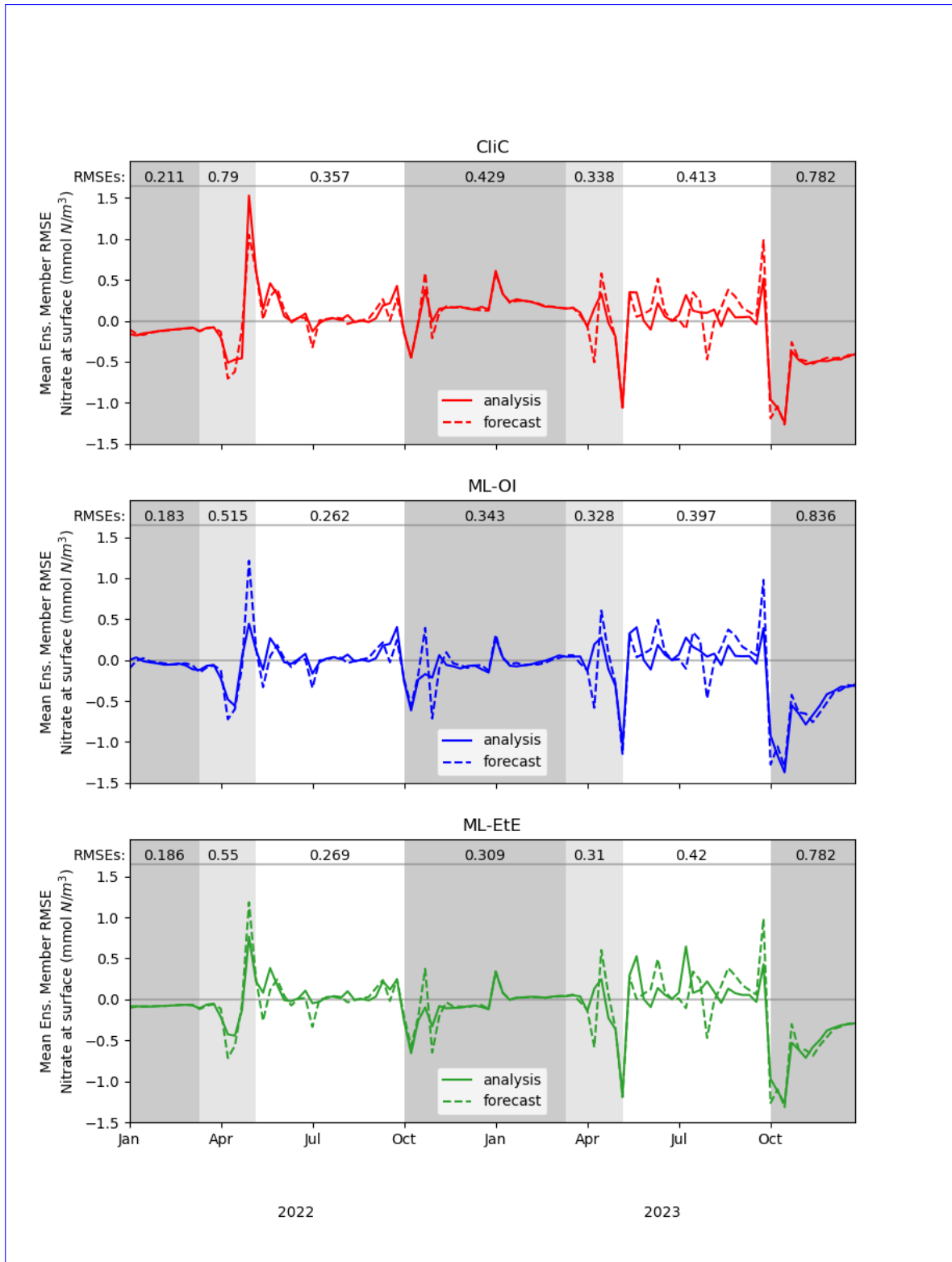


Figure 5: A comparison of mean nitrate background and analysis increments errors produced in the single-model run in runs for schemes using “online” cycled-DA at the surface. The first panel shows the analysis increments made during the climatological correlations CliC run RMSE (solid red) and the difference between the background state and the truth RMSE (dashed red) of the CliC runs. The second panel shows the analysis increments made by the ML-predicted correlations ML-OI run RMSE (solid blue) and the difference between the background state and the truth RMSE (dashed blue) of the ML-OI runs. The third panel shows the analysis increments predicted directly by the ML-EtE run RMSE (solid green) and the difference between the background state and the truth RMSE (dashed green) of the ML-EtE runs. In For each panel seasonal regime, the  $R$ -value represents the correlation between the mean analysis increments and RMSE across the difference between the background entire period and across all initialisations is given at the truth top of each panel. Shading indicates the system regimes previously outlined in Sect. 4.1: “light-limited” (white/dark grey), “bloom” (light grey) and “nutrient-limited” (dark grey/white).

663 These single-model schemes are then investigated further in Fig. 5, looking at the analysis incre-  
664 ments generated in the “online” setting, and their differences to the truth.

665 While Fig. 4 shows that they improve on average, Fig. 5 gives ~~some details~~ detail on when improve-  
666 ments are made. The runs shown here receive the same observations of the truth, and use the same  
667 initial conditions and forcing. However, the cycled “online” DA implies that the background state of a  
668 given time step will differ between methods. Nevertheless, we can see when ML-OI, or ML-EtE, make  
669 improvements over CliC. A clear example of this is the improved ~~predictions~~ estimations during the  
670 bloom period, where the ~~ML-predicted correlations provide a series of increments that are much closer~~  
671 ~~to the truth than the climatological increments~~ ML-estimated methods both provide a lower RMSE  
672 than the CliC. This shows that both ML methods are able to react to the timing of the bloom event  
673 much more accurately than climatology can. During the ~~nitrate-limited~~ nutrient-limited period, we  
674 generally see comparable performance across the ~~increments~~ methods, as the expected correlations are  
675 generally high. ~~However, the~~ and the ML-OI method can only weakly estimate the variation over this  
676 period (as seen previously in Fig. 3). ~~ML-EtE approach seems to more accurately capture the largest~~  
677 ~~analysis increments~~ provides no obvious advantage in this case, and explains why all methods struggle  
678 to make good increments in the second year of analyses (where the 7-day forecast errors are typically  
679 larger than in the first year). We can also see that each approach makes little to no adjustment during  
680 the majority of the light-limited regime. ~~The increments in~~ However, the increments by both ML-OI  
681 and CliC seem to introduce some noise around this time period, while the ML-EtE seems to more  
682 reasonably predict negligible increments at the start of this regime appear to yield a prolonged benefit  
683 once the system becomes inactive over winter. While these increments are unlikely to have any major  
684 impact on the system, it is interesting to note that the increments from ~~ML-EtE approach, which~~  
685 ~~replicates the end-to-end process of the EnKF increments, seem to more~~ each approach, accurately  
686 reflect the expected “decoupling” of total chlorophyll to nitrate at this time of year. This ~~suggests~~  
687 ~~that ML-EtE has successfully learnt the analysis increments~~ corroborates with offline experiments of  
688 Fig. 3, where the ML-OI model (and indeed, the climatological estimates) struggle to replicate the  
689 correlations over winter, as there is no strong dynamical relationship between the variables at this  
690 time.

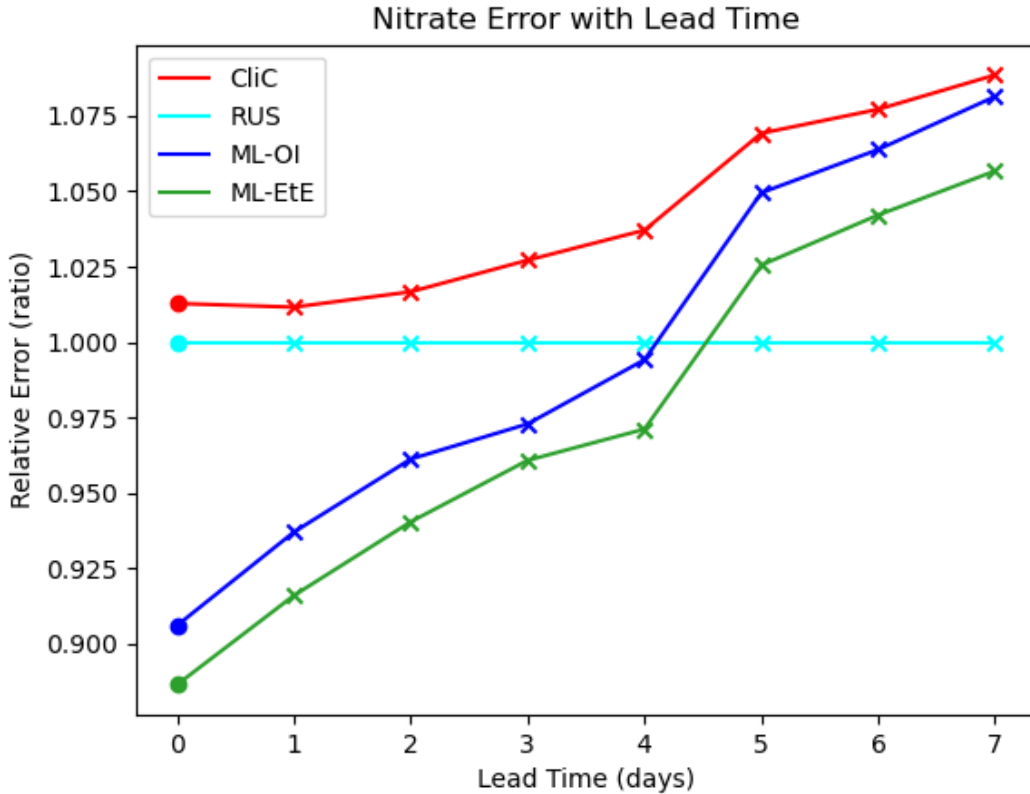


Figure 6: The forecast RMSE of each scheme relative to that of the RUS scheme at daily lead time [increments/intervals at the surface](#). For each scheme, the dot indicates the relative analysis error, while the crosses shows the relative forecast error for each day of lead time until a maximum lead time of 7 days - which is the total time between observations of total chlorophyll [in these experiments](#).

691 Finally, Fig. 6 shows analysis and forecast errors in nitrate in each scheme where errors are nor-  
692 malised against the error in the RUS scheme. The daily climatological correlations (CliC, red line)  
693 degrade the analysis error and then make worse forecasts at every lead time when compared to the  
694 RUS scheme, which does not update the nitrates at all. As previously noted, both ML approaches  
695 provide an analysis state that is approximately 8-12% better than the not-updated RUS nitrate. Im-  
696 proved forecasts then persist for approximately 4-5 days of lead time, only reaching an increased  
697 relative error after 5 days. For all lead times, ML-EtE outperforms ML-OI. While this is a net benefit  
698 to the forecasts of the system, it highlights the difficulty with partially updating a highly non-linear  
699 system. In this, it is clear that each attempt to update the nitrate results in an eventual error growth  
700 beyond simply not updating the system. Part of this could stem from the role of nitrate as a limiting  
701 nutrient; in that it is either available to allow phytoplankton growth, or not. This means that when  
702 [predicting-estimating](#) an increment for nitrate, we can know that some nitrate should be present or  
703 not, but a precise, continuous quantity that should be added or removed is not information that can  
704 necessarily be inferred from the observation of total chlorophyll. However, this error growth could also  
705 result from the analysis increments introducing some additional imbalance in other quantities of the  
706 system that also need correcting, and the complex marine BGC processes are inter-dependent. These  
707 imbalances and forecast error growths are discussed further in the following Sect. 4.3, when updating  
708 the additional marine BGC variables.

709 In the context of operational systems, such as those implemented by the UK Met Office, total  
710 chlorophyll is assimilated on a daily cycle, and then a forecast is produced for up to six days of lead  
711 time from these improved initial conditions. These results imply that there are huge gains to be  
712 made not only in short term forecasting (before errors saturate again), but also in reanalysis products  
713 that assimilate data with higher frequency, as the ML approaches substantially outperform the RUS  
714 scheme at this point.

715 **4.3 Extending the set of updated variables**

716 In this section, we demonstrate the additional benefit of predicting-estimating updates not just for  
 717 nitrate, but for nearly all marine BGC variables. In Fig. 7, we compare the different ML approaches  
 718 for updating an extended set of unobserved marine BGC variables, as well as the previous system  
 719 that only updates nitrate.

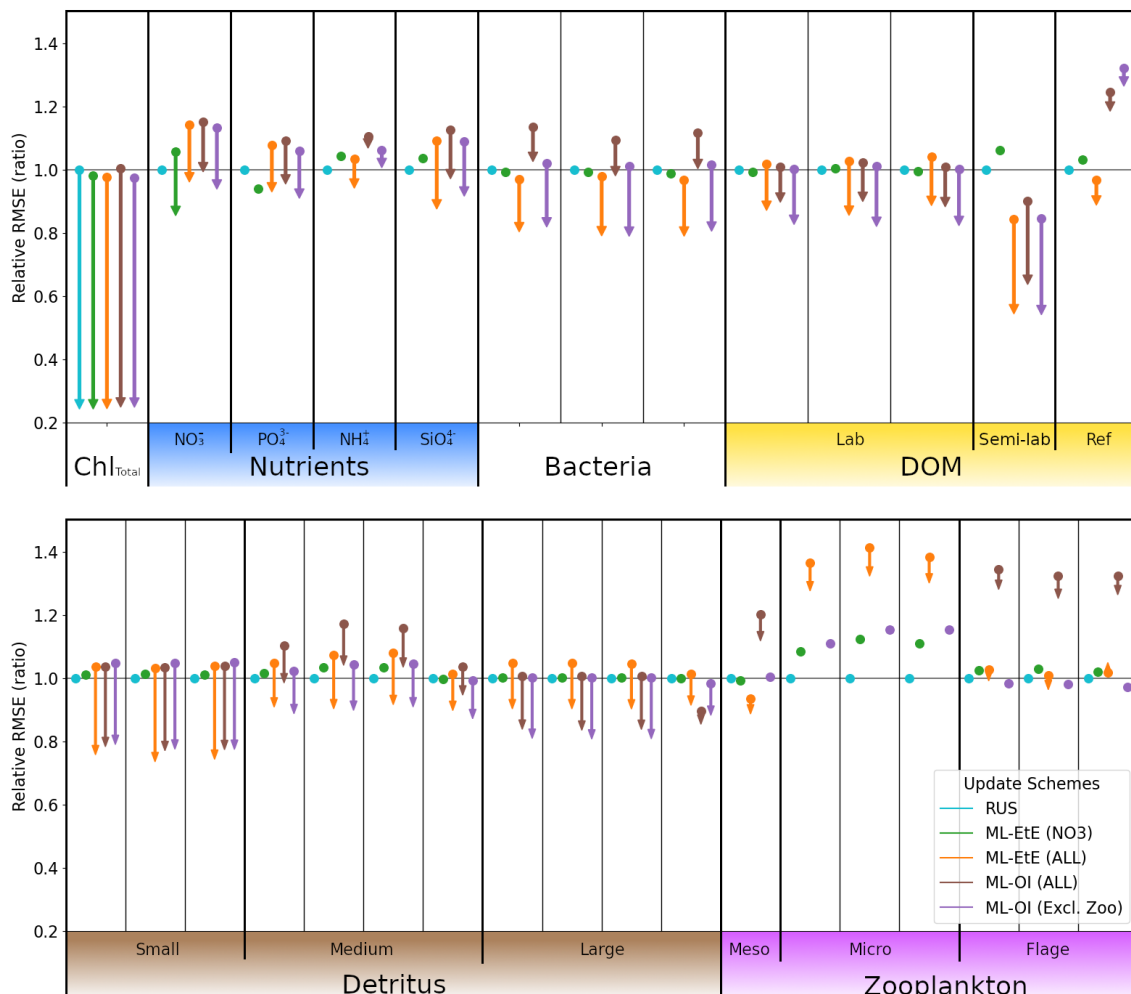


Figure 7: A comparison of the different schemes implemented to update the various components of the ERSEM BGC model at the L4 location. Surface RMSEs are calculated as an average across the entire online test period and across all initialisations. Forecast state (and equivalently background state) RMSEs at 7-day lead times are shown with dots, and the corresponding arrows indicate the analysis RMSEs (no arrow indicates the variable is not updated by the scheme). All RMSEs are relative to the RMSE in the RUS scheme shown in cyan, which only updates the total chlorophyll and PFTs as described in Sect. 2.4.1. The RMSEs of the ML-EtE (NO3) scheme, green, are from the same experiment shown previously in Sect. 4.2, and is used as another comparison point for the extended schemes. The ML-EtE (ALL) and ML-OI (ALL), orange and brown respectively, extend the ML schemes described in Sects. 3.2 and 3.1 to update all other pelagic variables. Finally, ML-OI (Excl. Zoo) (purple) updates all pelagic variables, excluding the zooplankton types. The chemical components of each variable class/type follow the same order (left to right) as Table 1.

720 The RUS scheme, described in Sect. 2.4.1, is used as a benchmark for the extended schemes, and  
 721 so values shown in Fig. 7 are RMSEs for 7-day forecasts relative to the RMSE of the RUS method  
 722 (averaged across the 104 forecast-analysis cycles of the entire test period). Again, we recall that the  
 723 RUS does not update any variables beyond total chlorophyll (shown) and its constituent PFTs (not  
 724 shown). The ML-EtE (NO3) scheme (green), which updates only nitrate, is carried over from the  
 725 previous section (as it performed best), to act as another point of comparison for the extended schemes.  
 726 Before discussing the extended schemes, we can see from Fig. 7 the dynamical impact that the updates  
 727 of ML-EtE (NO3) have on other (i.e. non-updated) marine BGC variables in the system. Generally,

728 the change in RMSE for these non-updated variables is very small, with the largest improvement being  
729 to phosphate and the largest degradation to zooplankton types – particularly microzooplankton. It  
730 also slightly degrades the ammonium and silicate concentrations that are not updated during the  
731 analysis. While this shows that updating a key nutrient, such as nitrate, can have wider impact on  
732 the system through dynamical adjustment, the generally beneficial results of the extended schemes  
733 (discussed below) point towards needing a DA system that can make reasonable adjustments to a  
734 wider set of marine BGC variables.

735 Our next scheme, ML-EtE (ALL) (orange), again follows the approach described in Sect. 3.2, but  
736 extends updates to all shown pelagic variables by predicting-estimating analysis increments directly  
737 from each background state and total chlorophyll increment. In this L4 setup, this is generally  
738 the best performing scheme, improving unobserved forecast and analysis RMSEs by between 10 –  
739 50%. We emphasise that while many of the 7-day forecast RMSEs are similar or even degraded  
740 compared with RUS, many of the benefits from an improved analysis persist over a significant portion  
741 of the forecast window (as is shown and discussed later). The most notable exceptions are the  
742 zooplankton which, despite having analysis increments in the correct direction, still return noticeably  
743 higher forecast/analysis RMSEs than most other schemes. Zooplankton have more interactions with  
744 other system components, existing at a higher trophic level, which result in a wider range of uncertainty  
745 for their behaviour. This also suggests they have generally weaker correlations with total chlorophyll.  
746 In our configuration, the RMSEs of the zooplankton group are clearly highly sensitive to updates in  
747 other variables, whether only nitrate is assimilated or a broader set is considered. When correlations  
748 between total chlorophyll and zooplankton are weak (as in Fig. A.2), chlorophyll increments contribute  
749 little information for correcting the zooplankton field, which is already strongly influenced by changes  
750 elsewhere in the system.

751 The ML-OI (ALL) scheme (brown) described in 3.1, extends updates to all shown pelagic variables  
752 by predicting-estimating the inter-variable correlation from the background state only. This prediction  
753 estimation is then combined with daily varying climatological variances to update the marine BGC  
754 state. This method is also shown to be somewhat effective, generally providing similar behaviour to  
755 the ML-EtE (ALL) scheme. ~~It also suffers~~, or at least reducing the RMSE from forecast to analysis  
756 (even if it is still worse than the RUS in some cases). It does suffer from the same difficulty in predicting  
757 estimating zooplankton updates, to an even greater degree, which causes some further imbalance in  
758 the system. This becomes clear when we exclude the zooplankton types from the updating in the  
759 ML-OI (Excl. Zoo) approach (purple), since it generally equals or makes small improvements over  
760 the ML-EtE (ALL) and ML-OI (ALL) schemes.

761 Figure. A.2 shows that correlations between surface total chlorophyll and silicate (or phosphate) are  
762 as strong as those with nitrate. This suggests that updates of similar magnitude might be expected for  
763 the other nutrients as well. However, it is not straightforward to infer how the system would evolve  
764 when all nutrients are updated simultaneously, based solely on the behaviour observed when only  
765 nitrate is updated. For example, Figure 7 reveals clear differences in the forecast and analysis RMSE  
766 for nitrate between the two ML-EtE configurations: one updating only nitrate and the other updating  
767 all nutrients. In the case for ML-EtE (ALL), updating all nutrients degrades the nitrate RMSE in  
768 both forecast and provides negligible impact on the analysis, while improving the analysis RMSEs for  
769 phosphate, ammonium and silicate. This outcome is notable, as it may point to assimilation biases  
770 introduced by the choice of update strategy.

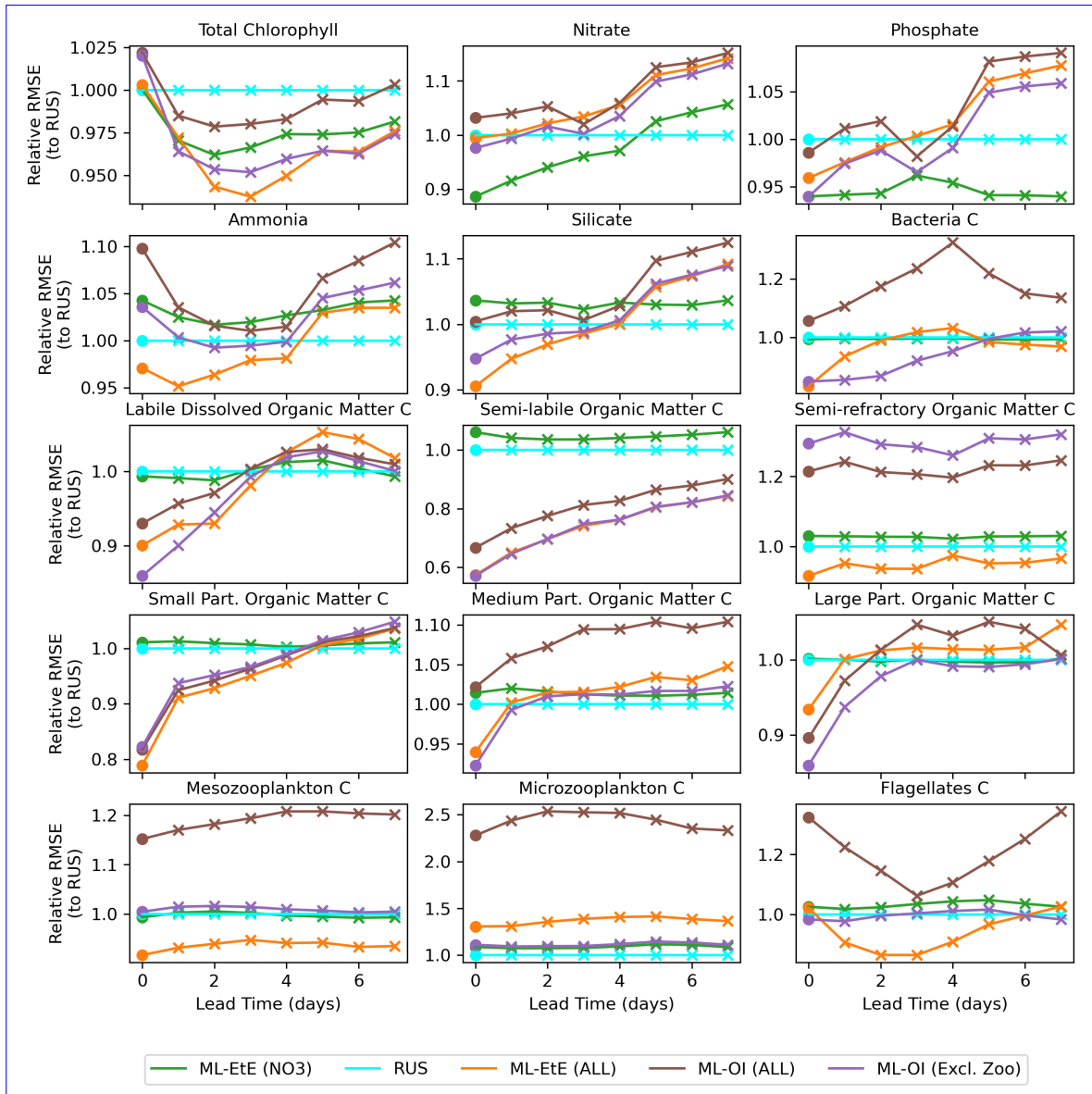


Figure 8: A comparison of the different schemes implemented to update the various components of the ERSEM BGC model at the L4 location (colours used are identical to Fig. 7). The forecast RMSE of each scheme relative to that of the RUS scheme at daily lead time intervals. For each scheme, the dots indicate the relative analysis error, while the crosses show the relative forecast error for each day of lead time until a maximum lead time of 7 days - which is the total time between observations of total chlorophyll.

771 Figure 8 shows the mean RMSE at multiple forecast lead times, for each method relative to the  
 772 RUS scheme. A variety of variables have been selected from Fig. 7 to represent the different behaviours  
 773 observed across each variable group. For many variables, there are notable improvements made over  
 774 the first 3 - 5 forecast lead times (e.g., most schemes for phosphate, ML-EtE (ALL) and ML-OI (Excl.  
 775 Zoo) in silicate, and the particulate organic matter variables), even if the gains do not persist for the  
 776 entire forecast window of 7 days. It is also important to note that the rate at which gains are lost  
 777 isn't necessarily quasi-linear for some variables. For example, small particulate organic matter loses  
 778 around half of its 20% gain from the analysis time to 1 day forecast time. Interestingly, the total  
 779 chlorophyll shows similar error relative to the RUS scheme for each other scheme at the analysis time  
 780 (they all handle the observations in the same way, so this should be expected), but each scheme then  
 781 makes improvements at essentially every other forecast lead time. This is likely due to the dynamical  
 782 adjustment of the model (where gains have been made in other variables) propagating through to the  
 783 total chlorophyll values as the model evolves. It is through this dynamical complexity, however, that  
 784 we see other variables do not improve or degrade monotonically as lead time increases (e.g., bacteria  
 785 and flagellates).

786 In summary the experiments from ~~Fig~~Figs. 7 and 8 show that the impact of DA analysis updates  
787 on the model forecast is not straightforward due to the ~~highly~~ non-linear and complex nature of the  
788 BGC model. Although in general it is true that increasing the number of updated variables benefits  
789 the forecasts, especially at shorter lead times (e.g., less than 5 days), this is definitely not true for  
790 every variable. ~~Furthermore, even if forecasts are improved during most of the~~ Some variables, for  
791 instance, show similar or worse forecast RMSE at a 7-day period relative to the RUSmodel (as can  
792 be anticipated based on Fig. 6), ~~around the 7-day lead time it does not really outperform the RUS~~  
793 ~~approach (based on the forecast RMSEs in Fig. 7).~~ lead time than the RUS, which highlights the  
794 need for specific results on each update strategy.

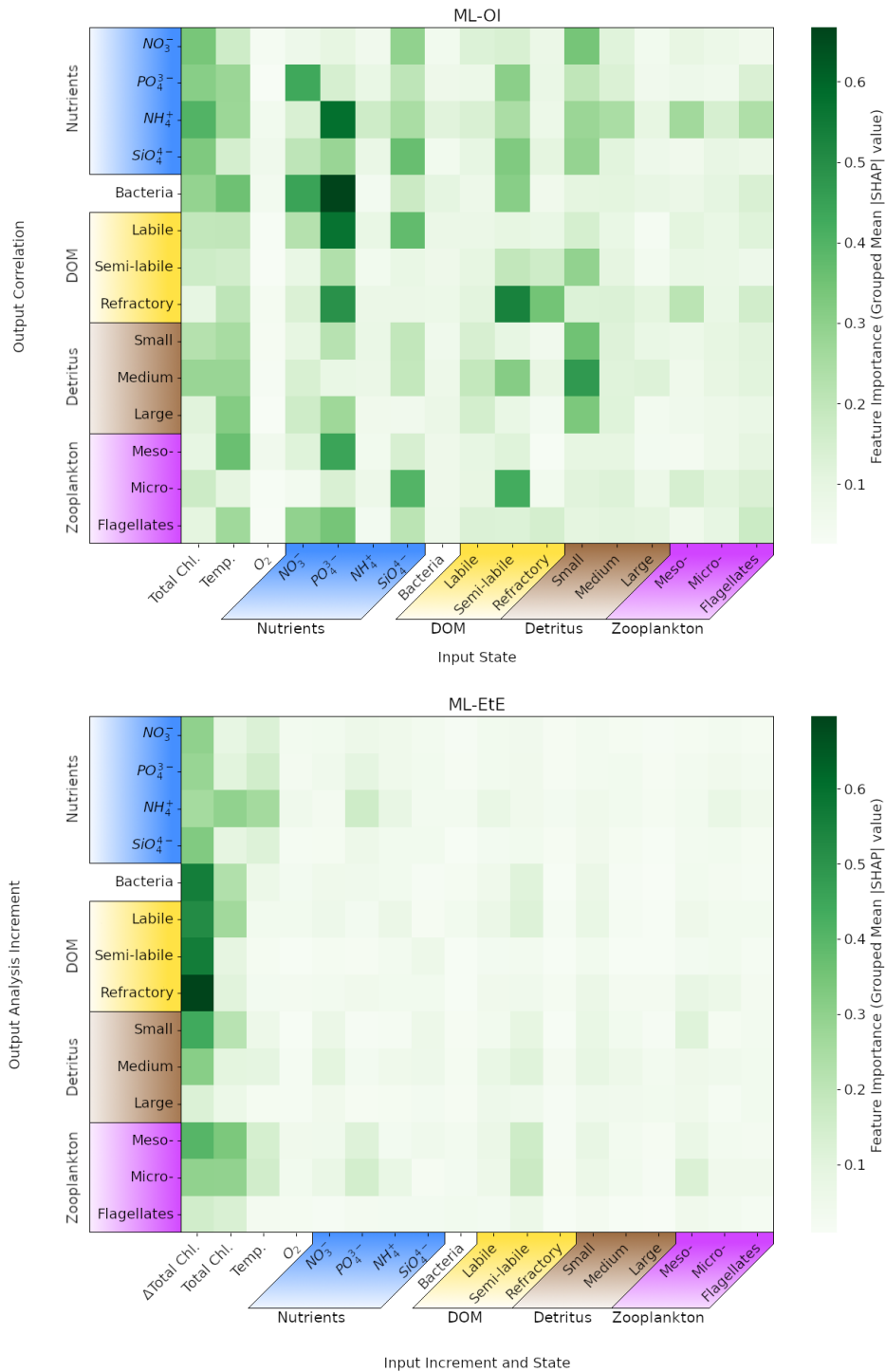


Figure 9: Mean absolute Shapley values estimated for the ML-OI (ALL) and ML-EtE (ALL), across their respective training datasets (see Sect. 3). Chemical components that belong to the same class or type (e.g., the carbon, nitrogen and phosphorus of bacteria) have been grouped as they are highly correlated. The variable names and chemical components are detailed in Table 1. The upper panel shows values for the extended ML-OI (ALL) model, and the lower panel for the extended ML-EtE (ALL) model. The grouped input features for each ML model are given on the  $x$ -axis, which make up the surface state for each pelagic variable. ML-EtE (ALL) has one additional feature,  $\Delta$ Total Chl., which represents the total chlorophyll analysis increment. The grouped output targets for each ML model are given on the  $y$ -axis, which correspond to the correlations [between total chlorophyll and each unobserved variable](#) for ML-OI (ALL), and the analysis increments for ML-EtE (ALL).

795 In Fig. 9, we interrogate the ML models using Shapley values (Sect. [??3.4.2](#)) to identify important  
796 ML-model features that are key to making accurate [predictionsestimations](#), and drive the connections

797 between observed total chlorophyll and unobserved variables. Specifically, Such a Shapley analysis  
798 also has the potential to help reduce the number of features needed for training future models (though  
799 this was not the aim of our experiments).

800 Fig. 9 shows the grouped mean absolute Shapley values for both the extended ML-OI (upper panel)  
801 and ML-EtE (lower panel) approaches. These are grouped as the separate chemical components of  
802 any class/type, and the resulting Shapley values, are very highly correlated. It is also important to  
803 note that Shapley values differ from a pure correlation between the input and output variables. This  
804 is because they capture both direct and interaction effects, account for non-linear relationships, and  
805 can explain a model’s decision-making rather than just measuring statistical association.

806 The ML-OI (ALL) Shapley values indicate that a broad range of input variables are important to  
807 the prediction-estimation of total chlorophyll correlations with unobserved variables, and highlight the  
808 general complexity of these interactions. We see that the state of temperature and total chlorophyll  
809 are moderately important across a broad set of variable groups. This makes sense as this ML model  
810 is predicting-estimating the correlation of a given variable with total chlorophyll, which is generally  
811 dependent on the state of total chlorophyll. However, this also implies that the seasonal regimes play a  
812 significant role in the predictionsestimations, as temperature is a clear identifier for the current time in  
813 the seasonal cycle. We note that the seasonal signal of other variables could also be important for the  
814 prediction-estimation of correlations as, in some cases, we see that at least one of the state variables  
815 in a group can be important to predicting-estimating the correlation between total chlorophyll and  
816 a state variable of the same group. For example, the state of small detritus is highly important to  
817 the entire group of detritus correlations, the states of some nutrients are generally important to the  
818 prediction-estimation of nutrients, and the semi-labile DOM is somewhat important to the wider DOM  
819 correlations. Some input features show no strong importance to any output targets. In particular, the  
820 zooplankton types seem largely unimportant in predicting-estimation their own correlation with total  
821 chlorophyll and the variables with a stronger signal have no obvious direct relationship. This may  
822 partially explain why zooplankton performs poorly when they are updated by the ML-DA schemes, as  
823 seen previously in Fig. 7, and points towards the difficulty and uncertainty associated with zooplankton  
824 in marine BGC modelling. This is further evidenced as the zooplankton types are unimportant as  
825 input features for all other correlation predictions-estimations as well. We also see that oxygen is  
826 largely unimportant to the prediction-estimation of the correlations. This observation is consistent  
827 with the known weak impact of oxygen assimilation in ERSEM on other modelled variables (Skákala  
828 et al., 2021). This would imply that both zooplankton and oxygen could be removed from the input  
829 feature set with little impact on the overall model performance.

830 The ML-EtE (ALL) Shapley values take on a distinctly different structure to those of the ML-OI  
831 (ALL). Recall that ML-EtE (ALL) has a different target than ML-OI (ALL), as it emulates analysis  
832 increments directly. It also has an additional input feature, the analysis increment of total chlorophyll,  
833 which is readily available in both the training dataset and at run-time. The most striking difference  
834 is that the total chlorophyll analysis increment dominates the prediction-estimation importances,  
835 showing the highest mean absolute value in almost all predictionsestimations. This is to be expected,  
836 as the total chlorophyll increment contains information about the observation, observational error  
837 and background model covariance, which are all necessary components of the unobserved analysis  
838 increment as described in Eq. (1). This makes sense considering the seasonal variation of the model  
839 and that total chlorophyll represents this variation quite reliably according to the regimes discussed in  
840 Sect. 4.1. The state variable input features show much less importance in ML-EtE (ALL) than in ML-  
841 OI (ALL), but it is sometimes still non-zero. These non-zero values seem to correlate somewhat with  
842 the most important input features seen in the ML-OI (ALL) approach, even if they are significantly  
843 reduced overall, suggesting that the state still contributes to the inherent flow dependencies of the  
844 analysis increments.

#### 845 4.4 Generalisation of machine learned-correlations to an unseen location

846 In this section, we test the performance of the extended ML approaches from Sect. 4.3 in the CWEC  
847 (Fig. 1), which exhibits different marine BGC behaviour than the L4 training location. In Fig. 10,  
848 we assess the performance of these ML models according to their 7-day forecast and analysis RMSEs.  
849 We then compare some general differences between the climatology of the two locations in Fig. 11,  
850 and then, with reference to the Shapley values shown previously in Fig. 9, we shall discern why the  
851 ML model might struggle extrapolating to the new location.

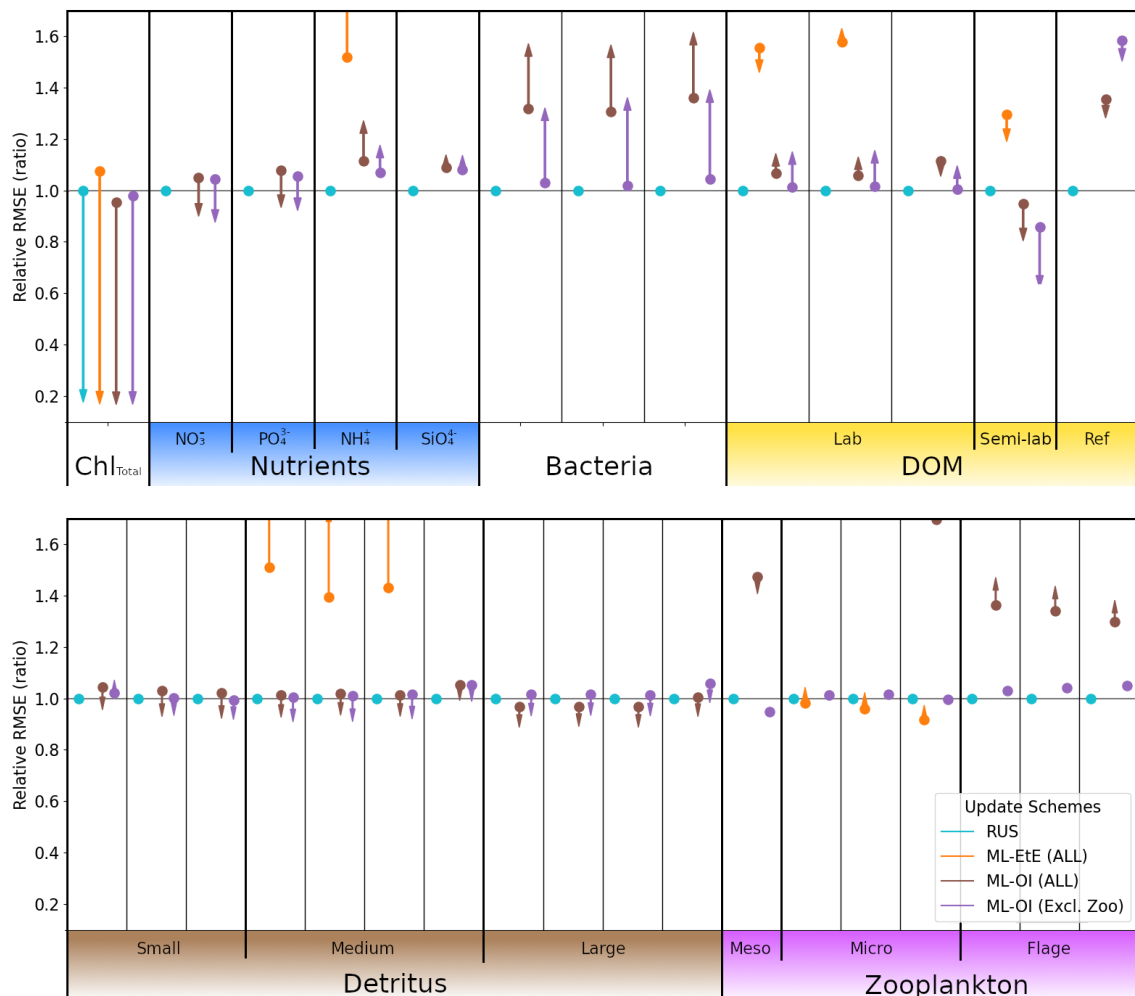


Figure 10: As Fig. 7, with the ML methods trained on L4, but applied to the new location CWEC. Dots and arrows that do not appear are off the scale.

852 Figure 10 again uses the RUS scheme as a comparison point for the ML model approaches, so  
 853 all RMSEs are given as a ratio of the RUS's RMSE value at the new location. The ML-EtE (ALL)  
 854 approach (orange) performs extremely poorly in this new location, with a large portion of the RMSEs  
 855 exceeding  $1.5\times$  the RUS background error (off the scale of Fig. 10). This is because the emulated  
 856 analysis increments of the EnKF at the L4 location fit the variability and scale of that (trained)  
 857 location and so, do not translate well to the new location. This means that, while the ML-EtE (ALL)  
 858 approach works well at the trained location (and fits the expected distribution of input data), in  
 859 practice its extendability to a new location is limited by both availability of training data and to the  
 860 new location's similarity to the original training location. The ML-OI (ALL) (brown) makes a marked  
 861 improvement over the ML-EtE (ALL) scheme, which is the reverse of the previous scenario at the L4  
 862 location. This is likely because the correlations ~~predicted-estimated~~ by the ML-OI scheme represent  
 863 a more location-agnostic relationship in the marine BGC variables, which can be used in combina-  
 864 tion with the climatological variances of CWEC to produce more location-appropriate increments -  
 865 ~~However~~(though this is not true for ammonium, bacteria and labile DOM, which produce a worse  
 866 analysis than forecast). Furthermore, this scheme still struggles to ~~predict-estimate~~ zooplankton cor-  
 867 relations, and so not updating the zooplankton as in the ML-OI (Excl. Zoo) scheme (purple) ~~produces~~  
 868 a broadly-improved result reduces the damage compared to ML-OI (ALL) - with any improvements  
 869 being marginal at best. In both ML-OI (ALL) and ML-OI (Excl. Zoo), we see that ~~the analysis for~~  
 870 detritus is generally improved relative to the RUS scheme (though these improvements are marginal,  
 871 and of similar magnitude to the worsened forecasts). Figure 11 shows that the climatological correla-  
 872 tions for these variables are generally similar in both locations (compare Fig. A.2 and Fig. A.3 to see  
 873 how these vary with time), with small detritus (originating largely from species with size  $< 20\mu m$ )  
 874 showing similarity in both climatological correlation and standard deviation. Since small detritus is  
 875 the most important input feature, in Fig. 9, for the ~~prediction-estimation~~ of detritus correlations in

876 the ML-OI models, it is reasonable to see why the improvements persist between the two locations.  
 877 We also see in Fig. 10 that both ML-OI models (brown and purple) make improvements to the analy-  
 878 sis RMSEs of nitrate, phosphorus and semi-labile DOM, which show relatively similar climatological  
 879 behaviour to L4 in Fig. 11 and Fig. A.1.

880 As all training is performed at one location, it is easy to hypothesise that the ML models have  
 881 overfitted to L4, specifically with regards to their use at other locations. Here, L4 is coastal and the  
 882 CWEC a more open area of ocean. This does not rule out the possibility that ML models trained on  
 883 a limited number of locations could extend their ~~predictions~~ estimations to spatial locations beyond  
 884 their set of training locations. However, it indicates that sparse training locations would need to be  
 885 chosen carefully, to appropriately cover the spread of behaviour in the system.

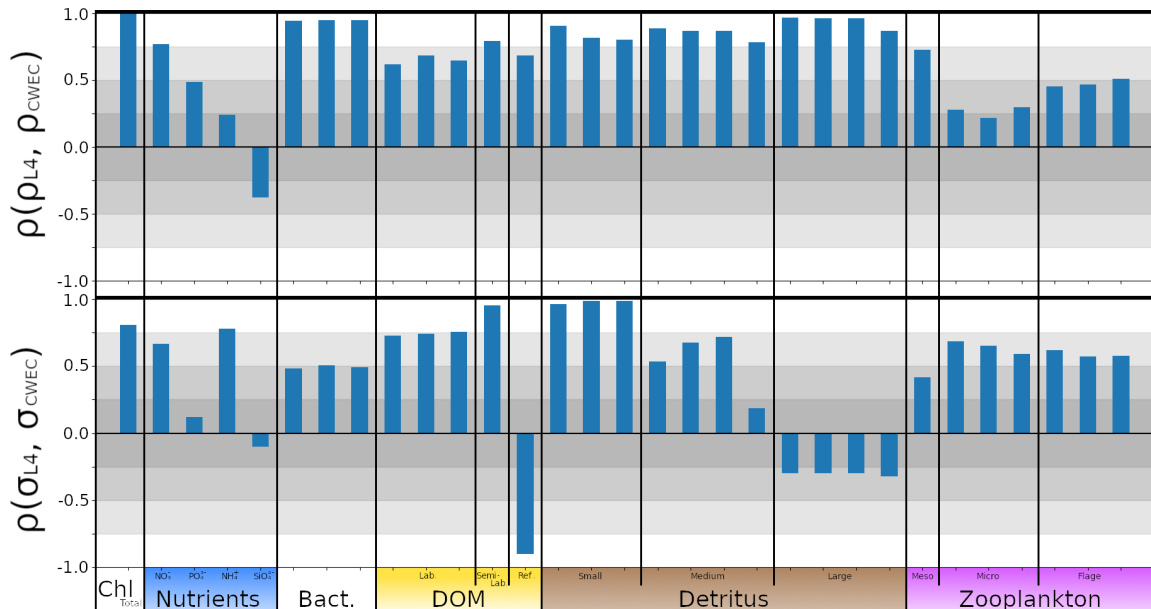


Figure 11: The top panel shows the correlation between a given variable’s climatological correlation signal at L4 and the CWEC,  $\rho(\rho_{L4}, \rho_{CWEC})$ . The bottom panel shows the correlation between a given variable’s climatological standard deviation signal at L4 and the CWEC,  $\rho(\sigma_{L4}, \sigma_{CWEC})$ . High correlations indicate that the model is behaving similarly in each location.

## 886 4.5 Viewpoint on scaling multivariate data assimilation to 3D models

887 We have ~~successfully~~ shown that ML methods can make improvements to the DA schemes of marine  
 888 BGC models when coupled to a 1D physical model - particularly in the shorter lead times of the  
 889 training location. The natural next question is how these results would scale when the marine BGC  
 890 model is coupled to a 3D physical model, such as NEMO (Nucleus for European Modelling of the  
 891 Ocean). A seemingly simple solution would be to run (once) a well-tuned, large EnKF, which can  
 892 then be used to train an ML model to be used operationally in an analogous 3D DA system that  
 893 presently updates only total chlorophyll. A reanalysis product with comprehensive statistics, or all  
 894 ensemble members available so statistics can be generated, would be ideal (Bonavita and Laloyaux,  
 895 2020; Brajard et al., 2021; Gregory et al., 2024). This circumvents the need to run an expensive  
 896 DA scheme operationally as the ML model could be trained offline, and then run significantly faster  
 897 while retaining the benefit of the statistics learned from a large ensemble. This would also allow the  
 898 analysis increments to be ~~predicted~~ estimated directly. However, state-of-the-art ensemble marine  
 899 BGC systems are still limited in scale and may not (yet) accurately represent the statistics needed for  
 900 multivariate DA (Skákala et al., 2024). Also, this approach would need to be repeated if/when the  
 901 observation network changes, which is likely given new observation missions and strategies (Telszewski  
 902 et al., 2018). A cheaper alternative would be to calculate the correlations in a free-run ensemble  
 903 dataset, as per the methods described in Sect. 3.1. This would be cheaper to create as there would be no  
 904 need to store and calculate both background states and analysis states. However, this approach cannot  
 905 calculate the analysis increments directly and instead must rely on the hybridisation of background  
 906 covariances/correlations into existing DA frameworks. Nevertheless our results on the 1D scenario  
 907 suggests that this is feasible and a good alternative to estimating the increments directly.

908 It is also worth considering how the data for these ML models should be sampled spatially in the  
909 3D case. Our results show there is some transferability between locations, as long as the dynamics  
910 are similar enough. In this, we suggest that a sparse forest of 1D models could be generated across  
911 the 3D domain, which aims to cover each region of sufficiently different biogeochemical behaviour.  
912 Previous work by Higgs et al. (2024) has split the North-West European Shelf into dynamically  
913 connected ecoregions, and this, or similar analysis, could be used as a guideline for generating these  
914 1D models. Furthermore, ML models could handle local multivariate aspects (a 0D transformation),  
915 while traditional DA methods (such as spatial correlations functions) manage 3D reconstruction (just  
916 as they manage the 1D reconstruction in our setup). A limitation of our two test locations is that  
917 they are not directly coupled, and could only be considered weakly coupled in the sense that their  
918 forcing data is extracted from the same 3D weather model. This could mean that 3D models have an  
919 advantage in locations having similar behaviour, as model grid points are much more likely to strongly  
920 correlate due to advection and ocean currents. However, the inverse could also be true, as the 1D  
921 models do not consider riverine input which can have substantial effects at the coast. Either way, the  
922 results suggest that some sparsity could be applied in extracting training data for these models, as  
923 long as each regime of BGC behaviour is represented in the selection. Introducing spatial variables  
924 like longitude and latitude could also improve the models ability to ~~predict-estimate~~  
925 correlations across the different horizontal locations.

926 An additional avenue worth exploring is whether training ML models on data from multiple,  
927 sufficiently different locations could improve their generalisability. Such an approach would allow for  
928 testing whether a more generalised model can (i) perform as well as a specialised model at its training  
929 locations, and (ii) transfer more successfully to new, unseen locations. While this is beyond the scope  
930 of the present study, it represents an interesting line of future work, particularly when combined  
931 with transfer learning strategies. For instance, a model trained at one location could be used as  
932 pre-conditioning for training at a new site, with the expectation that the pre-conditioned model would  
933 require less additional data to adapt effectively (e.g., Hu et al., 2016). Together, these directions  
934 highlight the importance of balancing specialisation and generalisation when scaling ML-assisted DA  
935 from idealized 1D configurations to realistic 3D systems.

## 936 5 Conclusions

937 Marine biogeochemistry (BGC) models aim to represent the complex BGC processes necessary to  
938 understand and forecast ecosystem behaviour. Data assimilation (DA) plays a crucial role in en-  
939 suring model trajectories remain closely aligned with real-world observations, along with the need  
940 for continuous improvement of numerical ~~predictions.~~ ~~Both estimations.~~ In this study, we used a  
941 synthetic “perfect model” setup as a necessary first step to explore ML-assisted DA under controlled  
942 conditions, but a natural direction for future work is to apply these approaches to real observations  
943 to assess their practical value. Nevertheless, both numerical modelling and DA are computationally  
944 expensive for marine BGC (dealing with great complexity and many variables), requiring well-tuned  
945 and accurately sampled statistics to be effective. These statistics are often poorly estimated in the  
946 undersized ensemble-based methods that are affordable operationally. In turn, this leads to the use  
947 of climatological ~~forecast-background~~ error covariance matrices in deterministic models, or simply not  
948 updating unobserved variables. This section concludes our work in relation to the research questions  
949 set out towards the end of Sect. 1 (reproduced below in italics).

950 *(a) Can we make improvements to the existing univariate scheme by updating a limited set of*  
951 *additional variables with an ML model to ~~predict-estimate~~ correlations or analysis increments?* In  
952 this study, we have demonstrated that neural networks can effectively learn statistical relationships  
953 between total chlorophyll (the only observed variable) and various pelagic BGC model variables.  
954 With machine learning (ML), we achieve ~~significant improvements over conventional approaches that~~  
955 ~~rely on climatological statistics or omit updates altogether~~ improvements over climatological statistics  
956 in the nitrate-only update framework. Our analysis of ~~ML-predicted-ML-estimated~~ nitrate updates  
957 illustrates that the ML methods behave in a largely coherent and meaningful manner, ~~reinforcing their~~  
958 ~~potential as an effective tool for improving BGC-DA.~~ While ML can degrade forecast skill in some  
959 unobserved variables compared to RUS or nitrate-only ML schemes, they nonetheless show promise,  
960 with their usefulness in more complex assimilation settings requiring further assessment.

961 *(b) Can these ML models be extended to effectively update all unobserved pelagic variables?* ML  
962 models can update almost all unobserved pelagic variables, supporting the broader applicability of ML  
963 in DA. ~~Some variables (notably zooplankton) do~~ In our configuration, zooplankton does not update  
964 well using either of the ML methods ~~that are extended to update~~ extended to all state variables

965 (~~namely~~ ML-OI (ALL) and ML-EtE (ALL)). ~~Zooplankton variables are~~, and is better treated in  
966 ~~these~~ hybrid DA schemes without being updated directly (as in ML-OI (Excl. Zoo)). However, this  
967 limitation may reflect the particular parameterisations of zooplankton-phytoplankton interactions,  
968 grazing, and mortality in the underlying BGC model, and may not generalise to other model setups.  
969 More broadly, we expect that variables less directly linked to or less sensitive to the observed quantity  
970 will be more difficult to update well. Since parameterisation choices can alter these relationships,  
971 new parameterisations would likely require retraining emulators, or alternatively, more flexible ML  
972 strategies such as transfer learning. Exploring such approaches in more diverse configurations remains  
973 an important avenue for follow-on investigations.

974 *(c) Is the ML model transferable to a new location after being trained on some other location?*

975 While a neural network trained in one water column exhibits partial transferability to other locations,  
976 challenges remain in fully generalising the model across spatial domains. This partial transferability  
977 is valuable, given the difficulty and cost of acquiring high-quality training data across large oceanic  
978 regions, and should be explored further in the context of 3D models. We discuss the feasibility of  
979 this, and propose a methodology for doing so. Future work should focus on refining transferability  
980 strategies, effective sampling strategy to allow for ergodic coverage ~~and~~ (i.e., ensuring statistical  
981 representativeness over space and/or time), and further evaluating the scalability of ML-driven DA  
982 in complex marine environments.

983 **A Characterisation of location biogeochemistry**

984 Figure A.1 shows that the CEWC is clearly less biologically productive than L4 with surface con-  
 985 centrations of total chlorophyll having a significantly lower median value, and a maximum that is  
 986 approximately 50% of L4's maximum. Each exhibit similar temperature values, as they are both lo-  
 987 cated within the English Channel. In the nutrients group, nitrate and phosphate values cover a similar  
 988 range in each location, but ammonium and silicate have little overlap. Bacteria and DOM concentra-  
 989 tions also show little similarity between locations. The small detritus concentrations are very similar  
 990 between both locations, but the medium and large detritus differ significantly, with CWEC covering  
 991 a much wider range of values than L4. Zooplankton concentrations also differ between the locations,  
 992 with CWEC producing much lower concentrations of zooplankton than the more biologically active  
 993 L4 location.

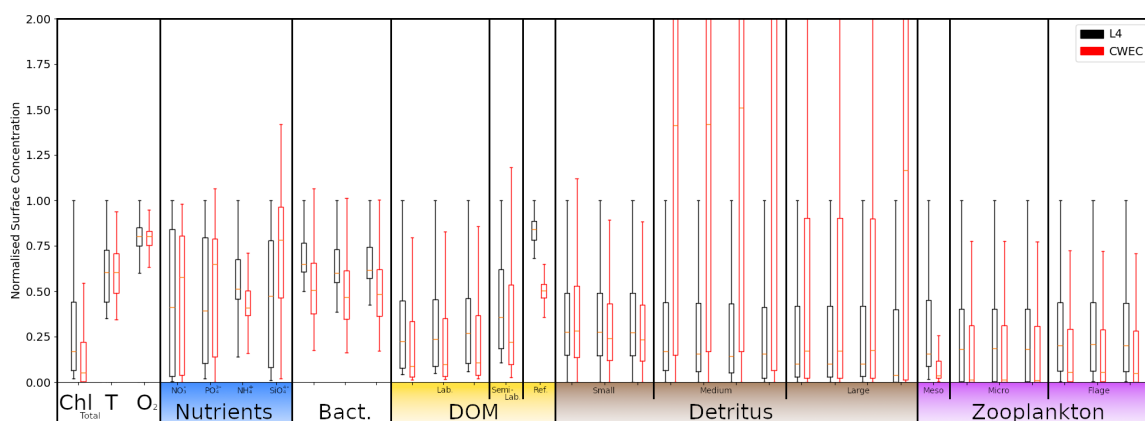


Figure A.1: Box and whisker plot showing the 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> percentile and upper and lower bound (excluding outliers larger than  $1.5 \times$  Inter quartile range) of each pelagic marine BGC variable for the RUS scheme in the online testing period. Values for L4 are given in black, and values for the CWEC are given in red. All values are normalised against the upper bound of L4. Chemical components are ordered according to Table 1. The label “T” corresponds to temperature.

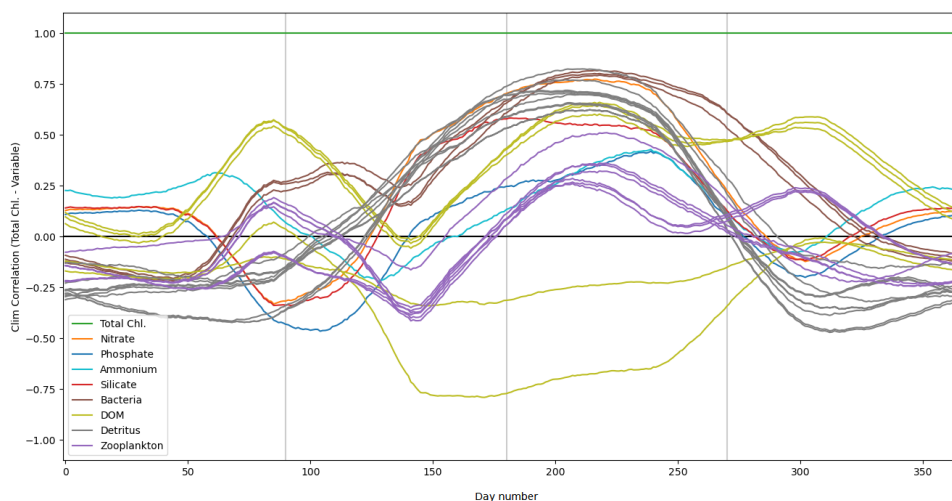


Figure A.2: Daily climatological correlations for each pelagic variable at the L4 location, calculated from the training free-run period of 2000-2014. Pelagic variables of the same class (according to Table 2) are shown with the same colour, except nutrients (nitrate, phosphate, ammonium and silicate) which are shown with separate colours.

994 The climatological correlations between total chlorophyll and other pelagic variables at the L4  
 995 location, shown in Fig. A.2, vary significantly according to the season. Variables of the same class

996 (see Table 1) generally exhibit very similar correlations. Correlations are much stronger during the  
 997 spring and summer months, as this period is more biologically active, and so the different model  
 998 components are going to be more closely coupled. Some variables, such as zooplankton, show a much  
 999 weaker correlative relationship with total chlorophyll.

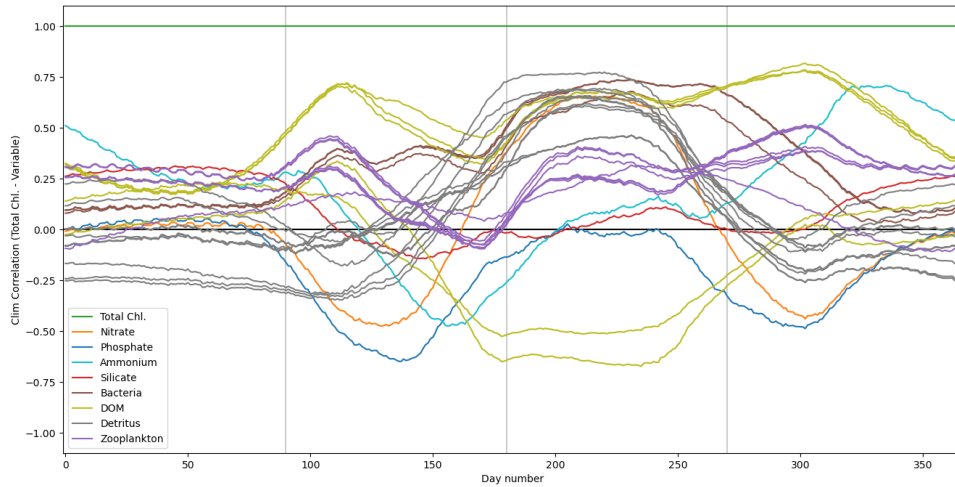


Figure A.3: As with Fig.A.2, except for the CWEC location from a free-run period of 2000-2010.

1000 Figure A.3 shows the climatological correlations between total chlorophyll and other pelagic vari-  
 1001 ables at the CWEC location. As with L4, the correlations of most variables show a much stronger  
 1002 correlation with total chlorophyll during the spring and summer, when the system is much more ac-  
 1003 tive. The correlations of nitrate are similar to those seen at the L4 location in Fig. A.2, following the  
 1004 pattern described Sect. 4.1. Zooplankton shows a weak correlation with total chlorophyll.

## 1005 **B Dynamical impact of updating only nitrate**

1006 Figure B.1 shows an extension of Fig. 4, with a representative set of variables that are unobserved  
 1007 and not updated (unlike nitrate which is also unobserved, but updated in this experiment). This  
 1008 clearly shows that the improvement of nitrate does not necessarily translate to an improvement in  
 1009 other variables, regardless of the method used to update the nitrate. This highlights the need for  
 1010 specific results of each update strategy.

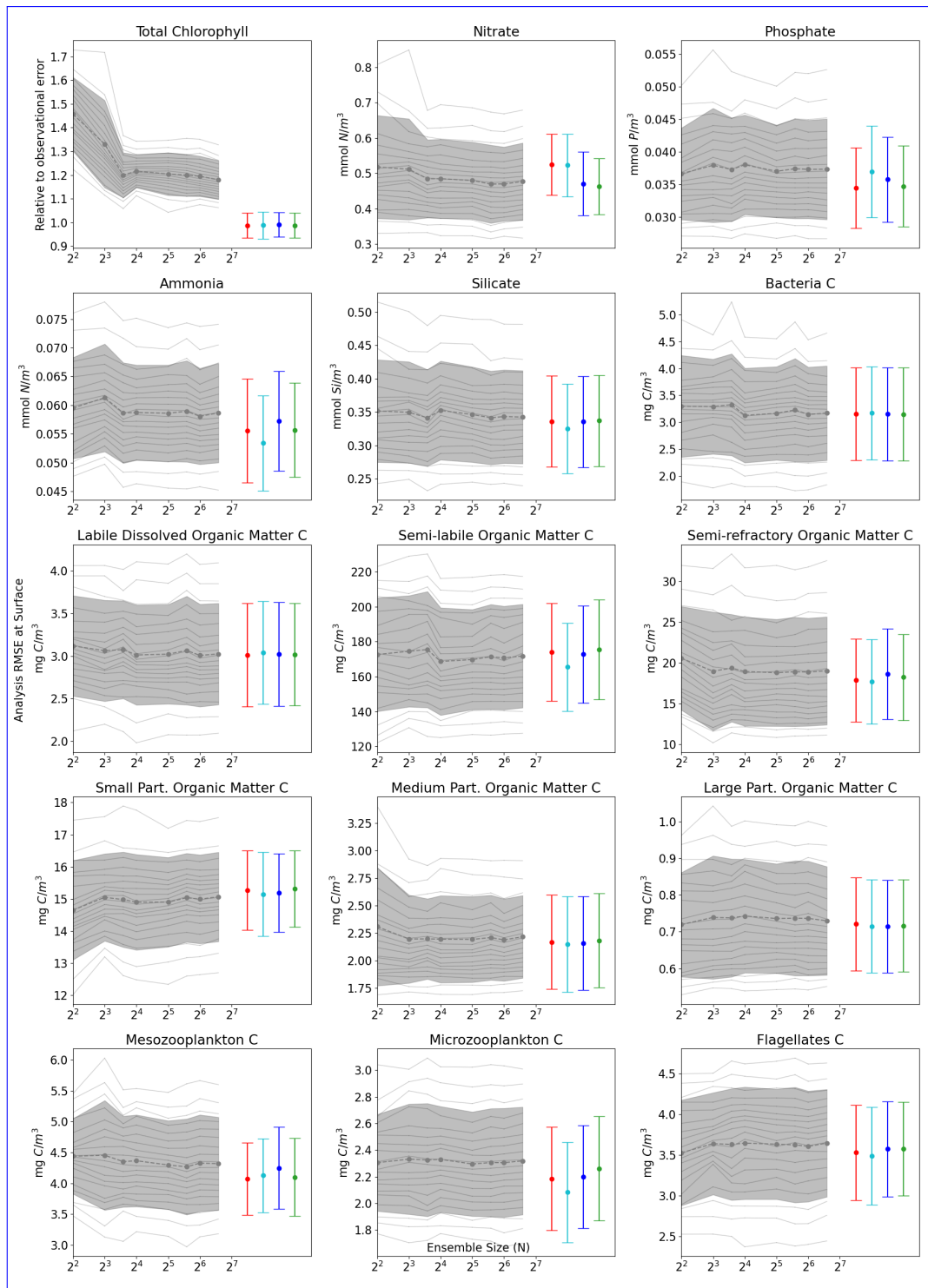


Figure B.1: An extension of Fig. 4, where panels for total chlorophyll (observed) and nitrate (unobserved, but updated) are the same. A representative set of additional variables (unobserved, not updated) are shown. A “C” indicates the panel is given for the carbon chemical component in the model.

1011 **Author Contributions**

1012 IH wrote and executed all code. All authors contributed to analysing and interpreting the results,  
 1013 proof-reading the manuscript, and adjusting the text.

1014 **Competing interests**

1015 At least one of the (co-)authors is a member of the editorial board of Biogeosciences. The peer-review  
1016 process was guided by an independent editor, and the authors also have no other competing interests  
1017 to declare.

## References

- 1018
- 1019 Anugerahanti, P., Kerimoglu, O., and Smith, S. L. (2021). Enhancing ocean biogeochemical models  
1020 with phytoplankton variable composition. *Frontiers in Marine Science*, 8:944.
- 1021 Artioli, Y., Blackford, J. C., Butenschön, M., Holt, J. T., Wakelin, S. L., Thomas, H., Borges, A. V.,  
1022 and Allen, J. I. (2012). The carbonate system in the North Sea: Sensitivity and model validation.  
1023 *Journal of Marine Systems*, 102:1–13.
- 1024 Asch, M., Bocquet, M., and Nodet, M. (2016). *Data assimilation: methods, algorithms, and applica-*  
1025 *tions*. SIAM.
- 1026 Baretta, J., Ebenhöf, W., and Ruardij, P. (1995). The European regional seas ecosystem model, a  
1027 complex marine ecosystem model. *Netherlands Journal of Sea Research*, 33(3-4):233–246.
- 1028 Baretta-Bekker, J., Baretta, J., and Ebenhöf, W. (1997). Microbial dynamics in the marine ecosystem  
1029 model ERSEM II with decoupled carbon assimilation and nutrient uptake. *Journal of Sea Research*,  
1030 38(3-4):195–211.
- 1031 Barth, A., Alvera-Azcárate, A., Licer, M., and Beckers, J.-M. (2020). DINCAE 1.0: A convolutional  
1032 neural network with error estimates to reconstruct sea surface temperature satellite observations.  
1033 *Geoscientific Model Development*, 13(3):1609–1622.
- 1034 Bertino, L., Ali, A., Carrasco, A., Lien, V., and Melsom, A. (2021). The Arctic Marine Forecasting  
1035 Center in the first Copernicus period. In *9th EuroGOOS International conference*, pages 256–263.
- 1036 Blackford, J. (1997). An analysis of benthic biological dynamics in a North Sea ecosystem model.  
1037 *Journal of Sea Research*, 38(3-4):213–230.
- 1038 Bocquet, M., Farchi, A., Finn, T. S., Durand, C., Cheng, S., Chen, Y., Pasmans, I., and Carrassi,  
1039 A. (2024). Accurate deep learning-based filtering for chaotic dynamics by identifying instabilities  
1040 without an ensemble. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(9).
- 1041 Bolding, K. and Villarreal, M. R. (1999). GOTM: A general ocean turbulence model: Theory, appli-  
1042 cations and test cases. Technical report, European Commission Tech. Rep. EUR 18745 EN.
- 1043 Bonavita, M. and Laloyaux, P. (2020). Machine learning for model error inference and correction.  
1044 *Journal of Advances in Modeling Earth Systems*, 12(12):e2020MS002232.
- 1045 Brajard, J., Carrassi, A., Bocquet, M., and Bertino, L. (2020). Combining data assimilation and  
1046 machine learning to emulate a dynamical model from sparse and noisy observations: A case study  
1047 with the Lorenz 96 model. *Journal of computational science*, 44:101171.
- 1048 Brajard, J., Carrassi, A., Bocquet, M., and Bertino, L. (2021). Combining data assimilation and  
1049 machine learning to infer unresolved scale parametrization. *Philosophical Transactions of the Royal*  
1050 *Society A*, 379(2194):20200086.
- 1051 Bruggeman, J. and Bolding, K. (2014). A general framework for aquatic biogeochemical models.  
1052 *Environmental modelling & software*, 61:249–265.
- 1053 Bruggeman, J., Bolding, K., Nerger, L., Teruzzi, A., Spada, S., Skakala, J., Ciavatta, S., et al.  
1054 (2024). Eat v1. 0.0: a 1d test bed for physical–biogeochemical data assimilation in natural waters.  
1055 *GEOSCIENTIFIC MODEL DEVELOPMENT*, 17(14):5619–5639.
- 1056 Buizza, C., Casas, C. Q., Nadler, P., Mack, J., Marrone, S., Titus, Z., Le Cornec, C., Heylen, E., Dur,  
1057 T., Ruiz, L. B., et al. (2022). Data learning: Integrating data assimilation and machine learning.  
1058 *Journal of Computational Science*, 58:101525.
- 1059 Burchard, H. (2009). Combined effects of wind, tide, and horizontal density gradients on stratification  
1060 in estuaries and coastal seas. *Journal of Physical Oceanography*, 39(9):2117–2136.
- 1061 Butenschön, M., Clark, J., Aldridge, J. N., Allen, J. I., Artioli, Y., Blackford, J., Bruggeman, J.,  
1062 Cazenave, P., Ciavatta, S., Kay, S., et al. (2016). ERSEM 15.06: a generic model for marine biogeo-  
1063 chemistry and the ecosystem dynamics of the lower trophic levels. *Geoscientific Model Development*,  
1064 9(4):1293–1339.

- 1065 Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. (2018). Data assimilation in the geosciences:  
 1066 An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*,  
 1067 9(5):e535.
- 1068 Cheng, S., Quilodr an-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss,  
 1069 B., Brajard, J., Xiao, D., Janjic, T., Ding, W., Guo, Y., Carrassi, A., Bocquet, M., and Arcucci,  
 1070 R. (2023). Machine learning with data assimilation and uncertainty quantification for dynamical  
 1071 systems: A review. *IEEE/CAA Journal of Automatica Sinica*, 10(6):1361–1387.
- 1072 Ciavatta, S., Brewin, R., Skakala, J., Polimene, L., de Mora, L., Artioli, Y., and Allen, J. I. (2018).  
 1073 Assimilation of ocean-color plankton functional types to improve marine ecosystem simulations.  
 1074 *Journal of Geophysical Research: Oceans*, 123(2):834–854.
- 1075 Ciavatta, S., Kay, S., Brewin, R. J., Cox, R., Di Cicco, A., Nencioli, F., Polimene, L., Sammartino, M.,  
 1076 Santoleri, R., Skakala, J., et al. (2019). Ecoregions in the Mediterranean Sea through the reanalysis  
 1077 of phytoplankton functional types and carbon fluxes. *Journal of Geophysical Research: Oceans*,  
 1078 124(10):6737–6759.
- 1079 Ciavatta, S., Kay, S., Saux-Picart, S., Butensch on, M., and Allen, J. (2016). Decadal reanalysis of  
 1080 biogeochemical indicators and fluxes in the North West European shelf-sea ecosystem. *Journal of*  
 1081 *Geophysical Research: Oceans*, 121(3):1824–1845.
- 1082 Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall’Olmo, G., Polimene, L., and Allen,  
 1083 J. I. (2014). Assimilation of remotely-sensed optical properties to improve marine biogeochemistry  
 1084 modelling. *Progress in Oceanography*, 127:74–95.
- 1085 Ciliberti, S. A., Gr egoire, M., Staneva, J., Palazov, A., Coppini, G., Lecci, R., Peneva, E., Matreata,  
 1086 M., Marinova, V., Masina, S., et al. (2021). Monitoring and forecasting the ocean state and bio-  
 1087 geochemical processes in the Black Sea: Recent developments in the Copernicus Marine Service.  
 1088 *Journal of Marine Science and Engineering*, 9(10):1146.
- 1089 Coppini, G., Clementi, E., Cossarini, G., Korres, G., Drudi, M., Amadio, C., Aydogdu, A., Agostini,  
 1090 P., Bolzon, G., Cret i, S., et al. (2021). The Copernicus marine service ocean forecasting system for  
 1091 the Mediterranean Sea. In *9th EuroGOOS International conference*, pages 272–279.
- 1092 Cossarini, G., Mariotti, L., Feudale, L., Mignot, A., Salon, S., Taillandier, V., Teruzzi, A., and  
 1093 d’Ortenzio, F. (2019). Towards operational 3D-Var assimilation of chlorophyll biogeochemical-Argo  
 1094 float data into a biogeochemical model of the Mediterranean Sea. *Ocean Modelling*, 133:112–128.
- 1095 Cossarini, G., Querin, S., Solidoro, C., Sannino, G., Lazzari, P., Di Biagio, V., and Bolzon, G. (2017).  
 1096 Development of BFMCOUPLER (v1. 0), the coupling scheme that links the MITgcm and BFM  
 1097 models for ocean biogeochemistry simulations. *Geoscientific Model Development*, 10(4):1423–1445.
- 1098 Council, N. R., on Geosciences, C., Science, W., Board, T., Board, O. S., on the Causes, C., and  
 1099 of Coastal Eutrophication, M. (2000). *Clean Coastal Waters: Understanding and Reducing the*  
 1100 *Effects of Nutrient Pollution*. National Academies Press.
- 1101 Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A. (2009). Ocean acidification: the other  
 1102 CO2 problem. *Annual review of marine science*, 1(1):169–192.
- 1103 Dowd, M., Jones, E., and Parslow, J. (2014). A statistical overview and perspectives on data assimi-  
 1104 lation for marine biogeochemical models. *Environmetrics*, 25(4):203–213.
- 1105 Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementa-  
 1106 tion. *Ocean dynamics*, 53:343–367.
- 1107 Fablet, R., Chapron, B., Drumetz, L., M emin, E., Pannekoucke, O., and Rousseau, F. (2021). Learning  
 1108 variational data assimilation models and solvers. *Journal of Advances in Modeling Earth Systems*,  
 1109 13(10):e2021MS002572.
- 1110 Falchetti, S., Conley, D. C., Brocchini, M., and Elgar, S. (2010). Nearshore bar migration and  
 1111 sediment-induced buoyancy effects. *Continental Shelf Research*, 30(2):226–238.
- 1112 Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., Crise, A., Edwards,  
 1113 C. A., Ford, D., Friedrichs, M. A., et al. (2019). Advancing marine biogeochemical and ecosystem  
 1114 reanalyses and forecasts as tools for monitoring and managing ecosystem health. *Frontiers in Marine*  
 1115 *Science*, 6:89.

- 1116 Fennel, K., Mattern, J. P., Doney, S. C., Bopp, L., Moore, A. M., Wang, B., and Yu, L. (2022). Ocean  
1117 biogeochemical modelling. *Nature Reviews Methods Primers*, 2(1):76.
- 1118 Fennel, K. and Testa, J. M. (2019). Biogeochemical controls on coastal hypoxia. *Annual Review of*  
1119 *Marine Science*, 11(1):105–130.
- 1120 Ford, D., Edwards, K., Lea, D., Barciela, R., Martin, M., and Demaria, J. (2012). Assimilating  
1121 GlobColour ocean colour data into a pre-operational physical-biogeochemical model. *Ocean Science*,  
1122 8(5):751–771.
- 1123 Ford, D., Key, S., McEwan, R., Totterdell, I., and Gehlen, M. (2018). Marine biogeochemical modelling  
1124 and data assimilation for operational forecasting, reanalysis, and climate research. *New Frontiers*  
1125 *in Operational Oceanography*, pages 625–652.
- 1126 Frölicher, T. L. and Laufkötter, C. (2018). Emerging risks from marine heat waves. *Nature commu-*  
1127 *nications*, 9(1):650.
- 1128 Galli, G., Wakelin, S., Harle, J., Holt, J., and Artioli, Y. (2024). Multi-model comparison of trends  
1129 and controls of near-bed oxygen concentration on the Northwest European Continental Shelf under  
1130 climate change. *Biogeosciences*, 21(8):2143–2158.
- 1131 Gehlen, M., Barciela, R., Bertino, L., Bresseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y.,  
1132 Ford, D., Lavoie, D., et al. (2015). Building the capacity for forecasting marine biogeochemistry  
1133 and ecosystems: recent advances and future developments. *Journal of Operational Oceanography*,  
1134 8(sup1):s168–s187.
- 1135 Geider, R., MacIntyre, H., and Kana, T. (1997). Dynamic model of phytoplankton growth and  
1136 acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light,  
1137 nutrient-limitation and temperature. *Marine Ecology Progress Series*, 148:187–200.
- 1138 Gobler, C. J. (2020). Climate change and harmful algal blooms: insights and perspective. *Harmful*  
1139 *algae*, 91:101731.
- 1140 Gregg, W. W. and Rousseaux, C. S. (2017). Simulating pace global ocean radiances. *Frontiers in*  
1141 *Marine Science*, 4:60.
- 1142 Gregory, W., Bushuk, M., Zhang, Y., Adcroft, A., and Zanna, L. (2024). Machine learning for  
1143 online sea ice bias correction within global ice-ocean simulations. *Geophysical Research Letters*,  
1144 51(3):e2023GL106776.
- 1145 Groom, S., Sathyendranath, S., Ban, Y., Bernard, S., Brewin, R., Brotas, V., Brockmann, C.,  
1146 Chauhan, P., Choi, J.-k., Chuprin, A., et al. (2019). Satellite ocean colour: Current status and  
1147 future perspective. *Frontiers in Marine Science*, 6:485.
- 1148 Gutknecht, E., Refray, G., Mignot, A., Dabrowski, T., and Sotillo, M. G. (2019). Modelling the marine  
1149 ecosystem of Iberia–Biscay–Ireland (IBI) European waters for CMEMS operational applications.  
1150 *Ocean Science*, 15(6):1489–1516.
- 1151 Hayashida, H., Steiner, N., Monahan, A., Galindo, V., Lizotte, M., and Lavoie, M. (2017). Impli-  
1152 cations of sea-ice biogeochemistry for oceanic production and emissions of dimethyl sulfide in the  
1153 Arctic. *Biogeosciences*, 14(12):3129–3155.
- 1154 Heinze, C. and Gehlen, M. (2013). Modeling ocean biogeochemical processes and the resulting tracer  
1155 distributions. In *International Geophysics*, volume 103, pages 667–694. Elsevier.
- 1156 Hemmings, J. C., Barciela, R. M., and Bell, M. J. (2008). Ocean color data assimilation with material  
1157 conservation for improving model estimates of air-sea CO<sub>2</sub> flux.
- 1158 Higgs, I., Skákala, J., Bannister, R., Carrassi, A., and Ciavatta, S. (2024). Investigating ecosystem  
1159 connections in the shelf sea environment using complex networks. *Biogeosciences*, 21(3):731–746.
- 1160 Hu, Q., Zhang, R., and Zhou, Y. (2016). Transfer learning for short-term wind speed prediction with  
1161 deep neural networks. *Renewable Energy*, 85:83–95.
- 1162 Jin, H., Song, Q., and Hu, X. (2019). Auto-keras: An efficient neural architecture search system.  
1163 In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data*  
1164 *mining*, pages 1946–1956.

- 1165 Jones, E. M., Baird, M. E., Mongin, M., Parslow, J., Skerratt, J., Lovell, J., Margvelashvili, N.,  
1166 Matear, R. J., Wild-Allen, K., Robson, B., et al. (2016). Use of remote-sensing reflectance to con-  
1167 strain a data assimilating marine biogeochemical model of the Great Barrier Reef. *Biogeosciences*,  
1168 13(23):6441–6469.
- 1169 Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects.  
1170 *Science*, 349(6245):255–260.
- 1171 Kerimoglu, O., Anugerahanti, P., and Smith, S. L. (2021). FABM-NflexPD 1.0: assessing an instantane-  
1172 ous acclimation approach for modeling phytoplankton growth. *Geoscientific Model Development*,  
1173 14(10):6025–6047.
- 1174 Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P., and Hoyer, S. (2021). Machine  
1175 learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sci-  
1176 ences*, 118(21):e2101784118.
- 1177 Le Traon, P. Y., Reppucci, A., Alvarez Fanjul, E., Aouf, L., Behrens, A., Belmonte, M., Bentamy,  
1178 A., Bertino, L., Brando, V. E., Kreiner, M. B., et al. (2019). From observation to information and  
1179 users: The copernicus marine service perspective. *Frontiers in Marine Science*, 6:234.
- 1180 Leeds, W., Wikle, C., Fiechter, J., Brown, J., and Milliff, R. (2013). Modeling 3-D spatio-temporal  
1181 biogeochemical processes with a forest of 1-D statistical emulators. *Environmetrics*, 24(1):1–12.
- 1182 Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances  
1183 in neural information processing systems*, 30.
- 1184 Mandal, S., Homma, H., Priyadarshi, A., Burchard, H., Smith, S. L., Wirtz, K. W., and Yamazaki,  
1185 H. (2016). A 1D physical–biological model of the impact of highly intermittent phytoplankton  
1186 distributions. *Journal of Plankton Research*, 38(4):964–976.
- 1187 Mattern, J. P., Fennel, K., and Dowd, M. (2012). Estimating time-dependent parameters for a  
1188 biological ocean model using an emulator approach. *Journal of Marine Systems*, 96:32–47.
- 1189 Mattern, J. P., Fennel, K., and Dowd, M. (2013). Sensitivity and uncertainty analysis of model hypoxia  
1190 estimates for the Texas-Louisiana shelf. *Journal of Geophysical Research: Oceans*, 118(3):1316–  
1191 1332.
- 1192 Mattern, J. P., Fennel, K., and Dowd, M. (2014). Periodic time-dependent parameters improving  
1193 forecasting abilities of biological ocean models. *Geophysical Research Letters*, 41(19):6848–6854.
- 1194 Mattern, J. P., Song, H., Edwards, C. A., Moore, A. M., and Fiechter, J. (2017). Data assimilation of  
1195 physical and chlorophyll a observations in the california current system using two biogeochemical  
1196 models. *Ocean Modelling*, 109:55–71.
- 1197 McEwan, R., Kay, S., and Ford, D. (2021). Quality information document for the CMEMS North  
1198 West European Shelf biogeochemical analysis and forecast. Technical report, CMEMS-NWS-QUID-  
1199 004-002 report.
- 1200 Nowack, P., Braesicke, P., Haigh, J., Abraham, N. L., Pyle, J., and Voulgarakis, A. (2018). Us-  
1201 ing machine learning to build temperature-based ozone parameterizations for climate sensitivity  
1202 simulations. *Environmental Research Letters*, 13(10):104016.
- 1203 Ouala, S., Fablet, R., Herzet, C., Chapron, B., Pascual, A., Collard, F., and Gaultier, L. (2018).  
1204 Neural network based Kalman filters for the spatio-temporal interpolation of satellite-derived sea  
1205 surface temperature. *Remote Sensing*, 10(12):1864.
- 1206 Pingree, R. and Griffiths, D. (1978). Tidal fronts on the shelf seas around the British Isles. *Journal  
1207 of Geophysical Research: Oceans*, 83(C9):4615–4622.
- 1208 Pradhan, H. K., Völker, C., Losa, S. N., Bracher, A., and Nerger, L. (2020). Global assimilation  
1209 of ocean-color data of phytoplankton functional types: Impact of different data sets. *Journal of  
1210 Geophysical Research: Oceans*, 125(2):e2019JC015586.
- 1211 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, f.  
1212 (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*,  
1213 566(7743):195–204.

- 1214 Sacco, M. A., Pulido, M., Ruiz, J. J., and Tandeo, P. (2024). On-line machine-learning forecast un-  
 1215 certainty estimation for sequential data assimilation. *Quarterly Journal of the Royal Meteorological*  
 1216 *Society*, 150(762):2937–2954.
- 1217 Sacco, M. A., Ruiz, J. J., Pulido, M., and Tandeo, P. (2022). Evaluation of machine learning tech-  
 1218 niques for forecast uncertainty quantification. *Quarterly Journal of the Royal Meteorological Society*,  
 1219 148(749):3470–3490.
- 1220 Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig,  
 1221 T., and Oschlies, A. (2017). Reviews and syntheses: parameter identification in marine planktonic  
 1222 ecosystem modelling. *Biogeosciences*, 14(6):1647–1701.
- 1223 Schmidtko, S., Stramma, L., and Visbeck, M. (2017). Decline in global oceanic oxygen content during  
 1224 the past five decades. *Nature*, 542(7641):335–339.
- 1225 Shulman, I., Frolov, S., Anderson, S., Penta, B., Gould, R., Sakalaukus, P., and Ladner, S. (2013). Im-  
 1226 pact of bio-optical data assimilation on short-term coupled physical, bio-optical model predictions.  
 1227 *Journal of Geophysical Research: Oceans*, 118(4):2215–2230.
- 1228 Simon, E. and Bertino, L. (2012). Gaussian anamorphosis extension of the DEKF for combined state  
 1229 parameter estimation: Application to a 1D ocean ecosystem model. *Journal of Marine Systems*,  
 1230 89(1):1–18.
- 1231 Simon, E., Samuelson, A., Bertino, L., and Mouysset, S. (2015). Experiences in multiyear combined  
 1232 state–parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using  
 1233 the ensemble Kalman filter. *Journal of Marine Systems*, 152:1–17.
- 1234 Skakala, J., Awty-Carroll, K., Menon, P. P., Wang, K., and Lessin, G. (2023). Future digital twins:  
 1235 emulating a highly complex marine biogeochemical model with machine learning to predict hypoxia.  
 1236 *Frontiers in Marine Science*, 10:1058837.
- 1237 Skakala, J., Bruggeman, J., Brewin, R. J., Ford, D. A., and Ciavatta, S. (2020). Improved represen-  
 1238 tation of underwater light field and its impact on ecosystem dynamics: A study in the North Sea.  
 1239 *Journal of Geophysical Research: Oceans*, 125(7):e2020JC016122.
- 1240 Skákala, J., Ford, D., Brewin, R. J., McEwan, R., Kay, S., Taylor, B., de Mora, L., and Ciavatta,  
 1241 S. (2018). The assimilation of phytoplankton functional types for operational forecasting in the  
 1242 Northwest European shelf. *Journal of Geophysical Research: Oceans*, 123(8):5230–5247.
- 1243 Skákala, J., Ford, D., Bruggeman, J., Hull, T., Kaiser, J., King, R. R., Loveday, B., Palmer, M. R.,  
 1244 Smyth, T., Williams, C. A., et al. (2021). Towards a multi-platform assimilative system for North  
 1245 Sea biogeochemistry. *Journal of Geophysical Research: Oceans*, 126(4):e2020JC016649.
- 1246 Skákala, J., Ford, D., Fowler, A., Lea, D., Martin, M. J., and Ciavatta, S. (2024). How uncertain and  
 1247 observable are marine ecosystem indicators in shelf seas? *Progress in Oceanography*, 224:103249.
- 1248 Smith, V. H. and Schindler, D. W. (2009). Eutrophication science: where do we go from here? *Trends*  
 1249 *in ecology & evolution*, 24(4):201–207.
- 1250 Smyth, T. J., Fishwick, J. R., Al-Moosawi, L., Cummings, D. G., Harris, C., Kitidis, V., Rees, A.,  
 1251 Martinez-Vicente, V., and Woodward, E. M. (2010). A broad spatio-temporal view of the western  
 1252 english channel observatory. *Journal of Plankton Research*, 32(5):585–601.
- 1253 Song, H., Edwards, C. A., Moore, A. M., and Fiechter, J. (2016). Data assimilation in a coupled  
 1254 physical–biogeochemical model of the California current system using an incremental lognormal 4-  
 1255 dimensional variational approach: Part 1—model formulation and biological data assimilation twin  
 1256 experiments. *Ocean Modelling*, 106:131–145.
- 1257 Sonnewald, M., Lguensat, R., Jones, D. C., Dueben, P. D., Brajard, J., and Balaji, V. (2021). Bridging  
 1258 observations, theory and numerical simulation of the ocean using machine learning. *Environmental*  
 1259 *Research Letters*, 16(7):073008.
- 1260 Sonntag, S. and Hense, I. (2011). Phytoplankton behavior affects ocean mixed layer dynamics through  
 1261 biological-physical feedback mechanisms. *Geophysical Research Letters*, 38(15).

- 1262 Telszewski, M., Palacz, A., and Fischer, A. (2018). Biogeochemical in situ observations—motivation,  
1263 status, and new frontiers. *New Frontiers in Operational Oceanography*, pages 131–160.
- 1264 Teruzzi, A., Bolzon, G., Feudale, L., and Cossarini, G. (2021). Deep chlorophyll maximum and  
1265 nutricline in the Mediterranean Sea: emerging properties from a multi-platform assimilated biogeo-  
1266 chemical model experiment. *Biogeosciences*, 18(23):6147–6166.
- 1267 Teruzzi, A., Dobricic, S., Solidoro, C., and Cossarini, G. (2014). A 3-D variational assimilation scheme  
1268 in coupled transport-biogeochemical models: Forecast of Mediterranean biogeochemical properties.  
1269 *Journal of Geophysical Research: Oceans*, 119(1):200–217.
- 1270 Umlauf, L. and Burchard, H. (2011). Diapycnal transport and mixing efficiency in stratified boundary  
1271 layers near sloping topography. *Journal of physical oceanography*, 41(2):329–345.
- 1272 Vagle, S., McNeil, C., and Steiner, N. (2010). Upper ocean bubble measurements from the NE Pacific  
1273 and estimates of their role in air-sea gas transfer of the weakly soluble gases nitrogen and oxygen.  
1274 *Journal of geophysical research: oceans*, 115(C12).
- 1275 van der Merwe, R., Leen, T. K., Lu, Z., Frolov, S., and Baptista, A. M. (2007). Fast neural network  
1276 surrogates for very high dimensional physics-based models in computational oceanography. *Neural*  
1277 *Networks*, 20(4):462–478.
- 1278 Wakelin, S. L., Artioli, Y., Butenschön, M., Allen, J. I., and Holt, J. T. (2015). Modelling the  
1279 combined impacts of climate change and direct anthropogenic drivers on the ecosystem of the  
1280 Northwest European continental shelf. *Journal of Marine Systems*, 152:51–63.
- 1281 Wakelin, S. L., Artioli, Y., Holt, J. T., Butenschön, M., and Blackford, J. (2020). Controls on near-  
1282 bed oxygen concentration on the Northwest European Continental Shelf under a potential future  
1283 climate scenario. *Progress in Oceanography*, 187:102400.