Progress in Oceanography 224 (2024) 103249

Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

How uncertain and observable are marine ecosystem indicators in shelf seas?

Jozef Skákala [a,b,*], David Ford [c], Alison Fowler [b,d], Dan Lea [c], Matthew J. Martin [c], Stefano Ciavatta [e]

[a] Plymouth Marine Laboratory, Plymouth, UK
[b] National Centre for Earth Observation, UK
[c] Met Office, Exeter, UK
[d] University of Reading, Meteorology Department, Reading, UK
[e] Mercator Ocean International, Toulouse, France

ARTICLE INFO

Keywords:
Marine ecosystem health indicators
Shelf seas
Uncertainty quantification
Model observability
Ensemble-variational data assimilation

ABSTRACT

Operational analysis and forecast products of shelf sea biogeochemistry often lack reliable information on uncertainty. This is problematic, as good quality uncertainty information is both requested by the product end-users and essential for data assimilation. To address this problem we developed a quality-assessed ensemble representation of many leading sources of uncertainty in a coupled marine physical-biogeochemical model of the North-West European Shelf. Based on these ensembles we have estimated the uncertainty of several marine ecosystem health indicators (MEHIs), acting as proxies for biological productivity, phytoplankton community structure, trophic fluxes, deoxygenation, acidification and carbon export. We have also evaluated how observable these MEHIs are from the most widely available observations of total chlorophyll (mostly from the surface), highlighting those MEHIs and locations that need to be better monitored. Our results show that the most uncertain and the least observable MEHI is the phytoplankton community composition, highlighting the value of its observations (and their assimilation) particularly in the UK regional waters. We demonstrate that daily operational estimates of the other MEHIs, produced by the Met Office, are fairly well constrained. We also quantify how much MEHI uncertainties are reduced when one substantially coarsens the MEHI spatial and temporal resolution, as in the global and/or climate applications.

1. Introduction

Shelf seas encompass only 7% of the surface of the world ocean, but play a disproportionately important role in the carbon cycle (20% of oceanic uptake of atmospheric carbon) and economy, providing home to the majority of global fisheries and aquaculture (Pauly et al., 2002; Borges et al., 2006; Jahnke, 2010). Reanalyses of marine health indicators have been vital to monitor the shelf sea ecosystem response to anthropogenic pressures and climate change (e.g. Fontana et al., 2009; Ciavatta et al., 2011, 2016, 2018). At the same time, forecasts of shelf sea ecosystem indicators (e.g. Skákala et al., 2018; Fennel et al., 2019) can provide warning signals to the end-users (e.g. industries, governments) about oceanic extreme events, such as hypoxia, eutrophication events, or potentially even harmful blooms.

In this study we focus on seven key marine ecosystem health indicators (MEHIs, for exact definitions see Table 1): (a) net primary production, (b) phytoplankton near-surface biomass, which can be used to extract phytoplankton phenology (we will further call it for simplicity "phenology"), (c) sea bottom oxygen, (d) particulate organic carbon (POC) in the deeper oceanic layer, (e) trophic efficiency, (f) phytoplankton community structure and (g) pH. The choice of these seven MEHIs was based on their relevance for monitoring a variety of threats and drivers: (i) deoxygenation/hypoxia (sea bottom oxygen), (ii) acidification (pH), (iii) eutrophication and biological productivity changes (net primary production, phenology, phytoplankton community structure), (iv) trophic web changes (phytoplankton community structure, trophic efficiency), and (v) impact on climate through carbon sequestration changes and carbon export (net primary production and the POC in deeper layers, e.g. Wilson et al., 2022). It should be noted that many of the MEHIs (e.g. oxygen, pH) are negatively impacted by the combination of anthropogenic pressures and climate change (Friedrich et al., 2012; Schmidtko et al., 2017; Kwiatkowski et al., 2020), which will make their monitoring in the future even more important.

* Corresponding author at: Plymouth Marine Laboratory, Plymouth, UK.
E-mail address: jos@pml.ac.uk (J. Skákala).

https://doi.org/10.1016/j.pocean.2024.103249
Received 21 September 2023; Received in revised form 8 March 2024; Accepted 24 March 2024
Available online 29 March 2024

Table 1
The MEHIs analysed in the study, their definition and relevance. Some of the MEHIs (trophic efficiency, bottom POC) could have been more naturally defined as fluxes rather than using the model state variables, however the adopted definitions have the advantage that they provide variables that could be in the future more easily measured and hence validated by the observations (Telszewski et al., 2018; Wang et al., 2019).

MEHI	Definition	Relevance
Phenology	Total phytoplankton carbon biomass integrated in the upper 5 m	Biological cycles
Net primary production	Net primary production integrated across the whole water column	Carbon seque- stration, biological productivity
Trophic efficiency	Carbon-ratio of total depth-integrated zooplankton to total depth-integrated phytoplankton	Trophic web changes, carbon fluxes
Phytoplankton community structure	Carbon-ratio of depth-integrated microphytoplankton (for ERSEM sum of diatoms and dinoflagellates) to total phytoplankton	Trophic web changes, harmful algae
POC at depth	Depth-integrated particulate organic carbon (POC) below the 500 m depth, or in the bottom 5% of the water column (wherever the bathymetry is shallower than 500 m)	Carbon export
Bottom oxygen	Oxygen averaged across the 10% of the water column nearest to the sea bottom	Hypoxia
pH	pH averaged across the water column up to the bottom, or the 500 m depth, if the bottom is deeper	Acidification

Although many MEHI products are being delivered by operational centres, time-dependent probabilistic uncertainties describing the "errors of the day" (Kalnay, 2003) are rarely provided. This is problematic, as any MEHI simulation is by its very nature only probabilistic, and one should therefore provide more information about the underlying probability density function (PDF) than a single value representing the PDF's mean, or median. It also leads to practical issues, as the lack of knowledge on uncertainty has been widely recognized as one of the main limitations on the end-user uptake of modelling-based products (e.g. Gehlen et al., 2015).

Furthermore, good-quality estimates of MEHI uncertainty are also essential to better estimate MEHIs themselves. This is because reanalyses and forecasts are based on systems that regularly (e.g. daily) constrain the coupled physical-biogeochemical models with the available observational data-sets, using data assimilation (DA, e.g. Bannister, 2017). Estimates of model forecast uncertainty are essential for DA, as they determine how much weight is given to the model versus the observations (together with the observation error estimates). The model forecast uncertainty estimates should ideally include flow-dependency, i.e. accurate description of how the uncertainty changes in time. But for variational operational systems, such as the system run at the Met Office for the North-West European Shelf seas (NWES, Mogensen et al., 2009, 2012), suitable flow-dependent uncertainty estimates are typically not available. Instead, parametrized and climatological uncertainties are used, sometimes originating from older, or research versions of the operational system. Although these climatological uncertainty estimates can be improved through iterative diagnostic methods (e.g. Hollingsworth and Lönnberg, 1986; Desroziers et al., 2005), insufficient flow-dependency can limit the ability of the system to estimate the true ocean state. This can sometimes result in overfitting the observations, as discussed in Skákala et al. (2022). This may be an issue for MEHI forecasts from the NWES operational system (Skákala et al., 2018, 2020, 2021, 2022), where the assimilation reduces substantial model biases (Skákala et al., 2018, 2022), but likely overfits observations, whose uncertainties and biases then contaminate the analysis (Skákala et al., 2022).

It should be noted that model uncertainties can have different natures, they can be of aleatory type, accounting for a process that is not deterministically described by the model, or can be of epistemic nature, e.g. corresponding to our lack of knowledge of model parameters, or initial values (Der Kiureghian and Ditlevsen, 2009; Hüllermeier and Waegeman, 2021). The aleatory uncertainties are irreducible (unless we improve the model), but unlike the epistemic uncertainties, they can be validated by comparing modelled PDFs with the distribution of events in nature. Since epistemic uncertainties and biases reflect our lack of knowledge, they can be reduced through incorporating observations of the system. However, when it comes to MEHI uncertainty, direct observations of many MEHIs are (despite recent progress, see Wang et al. (2019), Claustre et al. (2020), Brewin et al. (2021)) still rare (Telszewski et al., 2018). In such situations we typically have to rely on biogeochemical variables that are much easier to observe, but might be only indirectly related to MEHIs. The most commonly derived marine biogeochemical variable both from satellite ocean colour and in situ platforms (e.g. fluorescence measurements) is total chlorophyll-*a* concentrations (e.g. Brewin et al., 2021). We can then ask the question to what extent we observe a MEHI through observing the total chlorophyll-*a* variable. This question can be addressed through the concept of model (state) observability introduced in control theory by Kalman (1959, 1963).

We will use ensembles to estimate the uncertainty and observability of MEHIs produced by an assimilative system very similar to the one run operationally by the Met Office for NWES reanalyses and forecasts. The main difference between the DA system used here and the variational (3DVar) operational system is that the system here is 3D ensemble-variational (3DEnVar, following the terminology of Lorenc and Jardak, 2018) in the sense that we directly calculate (in a flow-dependent way) the background error variances and spatial covariances from the ensembles. The forecast version of the Met Office system assimilates ocean color (OC) satellite total chlorophyll-*a* and we will ask how observable each MEHI is from the available total chlorophyll data used in the assimilation. To address uncertainty and observability we will run two ensemble simulations: the uncertainty estimates will be based on an ensemble simulation assimilating physical data and total chlorophyll, similar to the Met Office operational forecasting and reanalysis systems. To estimate the impact of chlorophyll observations on MEHIs (i.e. the MEHI observability), we will compare it with a second ensemble simulation having the same set-up, but without the assimilation of total chlorophyll.

The paper is structured as follows: in the methodology part we describe the physical-biogeochemical model, the ensembles representing the different sources of uncertainties, the 3DEnVar DA system, the experiments from this study and the validation metrics and data. This will be followed by the results and discussion section, where we first validate the system using reliability diagnostics, and after that discuss

the impact of the new uncertainty estimates on the DA system skill. Then we analyse the different MEHI uncertainties and how they depend on spatial scale, discuss the MEHI observability, and finally, identify the MEHIs and locations with the greatest need to be observed.

2. Model, data and methods

2.1. The physical-biogeochemical model

The current operational Met Office system for NWES biogeochemistry is based on the physical model Nucleus for European Modelling of the Ocean (NEMO, Madec et al., 2015) coupled with the marine biogeochemistry European Regional Seas Ecosystem Model (ERSEM) (Baretta et al., 1995; Butenschön et al., 2016; Marine Systems Modelling Group, 2020), using the Framework for Aquatic Biogeochemical Models (FABM, Bruggeman and Bolding, 2014, 2020). The system assimilates data into the model using NEMOVAR DA software (Mogensen et al., 2009, 2012).

The physical model NEMO is a finite difference, hydrostatic, primitive equation ocean general circulation model (Madec et al., 2015). The NEMO configuration used in this study is similar to the one used by Ford et al. (2017), Skákala et al. (2018, 2020), and identical to the configuration used in Skákala et al. (2021, 2022): we use the CO6 NEMO version, based on NEMOv3.6, a development of the CO5 configuration explained in detail by O'Dea et al. (2017). The model has approximately 7 km spatial resolution on the Atlantic Margin Model (AMM7) domain using a terrain-following $z^*-\sigma$ coordinate system with 51 vertical levels (Siddorn and Furner, 2013). The lateral boundary conditions for physical variables at the Atlantic boundary were taken from the outputs of the Met Office operational $1/12°$ North Atlantic model (NATL12, Storkey et al., 2010); the Baltic boundary values were derived from a reanalysis produced by the Danish Meteorological Institute for CMEMS. We used river discharge based on data from Lenhart et al. (2010).

ERSEM is a lower trophic level ecosystem model that includes pelagic plankton, and benthic fauna (Blackford, 1997). The model splits phytoplankton into four functional types largely based on their size (Baretta et al., 1995): picophytoplankton, nanophytoplankton, diatoms and dinoflagellates. ERSEM uses variable stoichiometry for the simulated plankton groups (Geider et al., 1997; Baretta-Bekker et al., 1997) and each phytoplankton functional type (PFT) biomass is represented in terms of chlorophyll, carbon, nitrogen and phosphorus, with diatoms also represented by silicon. ERSEM predators are composed of three zooplankton types (mesozooplankton, microzooplankton and heterotrophic nanoflagellates), with organic material being decomposed by one functional type of heterotrophic bacteria (Butenschön et al., 2016). The ERSEM inorganic component consists of nutrients (nitrate, phosphate, silicate, ammonium and carbon) and dissolved oxygen. The carbonate system is also included in the model (Artioli et al., 2012).

In this work the physical (NEMO) model and the biogeochemical (ERSEM) model are two-way coupled, which means that not only does the physics drive marine biogeochemistry, but also marine biogeochemistry feeds back to physics through influencing the light attenuation and heat fluxes in the water-column. This is a development introduced recently in Skákala et al. (2022).

2.2. The different sources of uncertainty and the ensemble design

The purpose of this section is to offer a general discussion of the many sources of biases and uncertainties of coupled marine physical-biogeochemical models (e.g. Gehlen et al., 2015; Schartau et al., 2017; Payne et al., 2016; Garnier et al., 2016; Bonan and Doney, 2018; Meier et al., 2019; Fennel et al., 2022) and how these sources are represented in the specific case from this paper. The uncertainties are most typically modelled using Monte Carlo ensemble simulations, a computationally expensive method, but straightforward to implement (e.g. Dowd, 2007; Evensen, 2003). We identify three separate classes of sources for model uncertainty and bias: (i) uncertainty in model inputs, which include model forcing, lateral boundary conditions and initial value conditions, (ii) uncertainty in the values of the many model parameters, and (iii) structural deficiencies in the formulation of the model.

2.2.1. Model input uncertainties

Under model inputs we mean all the data that feed into the model to make it able to run, i.e. the initial value conditions, various types of model forcing (e.g. atmospheric, riverine), and the regional domain lateral boundary conditions. It is fairly straightforward to represent the uncertainties of the atmospheric forcing, as we have available atmospheric ERA5 ensemble reanalyses run by the European Centre for Medium-range Weather Forecasts (ECMWF, Palmer, 2019). The 10-member ERA5 ensemble has 31 km spatial resolution with outputs available at hourly resolution. This ensemble has been used here to force the NEMO-FABM-ERSEM model, representing a major part of the flux of uncertainty from the atmosphere into the ocean. This is a similar strategy to the one adopted in Lea et al. (2022). The biogeochemistry model also uses inputs of atmospheric carbon dioxide (CO_2) concentration and nitrogen deposition. These are not perturbed in this work but could be in future.

Representation of the riverine discharge uncertainties, which can be important on the NWES, has been already developed for the physical model (riverine runoff data, Zedler et al., 2023). To assess riverine impact on biogeochemistry these would need to be extended to a wide range of biogeochemical riverine inputs, such as nutrients, oxygen, or terrestrial dissolved organic matter (DOM). Neither the existing, nor new riverine perturbations were included here. The reasons for this are twofold: (i) Developing appropriate models of prior PDFs for the variety of biogeochemical tracer river discharges is a non-trivial task requiring deeper understanding of these uncertainties, their time-scales and cross-correlations among variables (e.g. a simple model based on scaling the perturbations of all the variables with respect to run-off perturbations might, or might not be sufficient). At the same time, using the existing runoff perturbations from Zedler et al. (2023), whilst avoiding perturbing the biogeochemistry riverine discharge data is not consistent either. (ii) Even if the riverine discharge perturbations were autocorrelated on long time-scales and correlated between rivers, their impact on significant parts of the NWES domain builds only relatively slowly in time, so any meaningful representations of these uncertainties would need long spin-up runs, far beyond what we could afford in this project.

The initial value uncertainties were partly represented through an initial model ensemble spin-up run (for details see Section 2.4) and through the observation perturbation scheme described in Section 2.3. Finally, the lateral boundary conditions were not perturbed in this work, but perturbations based on global ensemble runs will be considered in future.

2.2.2. Model parameter uncertainties

Marine models tend to have many free parameters, especially marine biogeochemistry models, which can be highly complex and poorly constrained (e.g. Schartau et al., 2017; Stefano et al., 2022; Fennel et al., 2022). For example, ERSEM has over 400 mostly phenomenological parameters with highly uncertain values (Butenschön et al., 2016).

Some of the parameters might be thought of as spatio-temporally varying. For example, even a complex biogeochemical model, such as ERSEM, cannot computationally afford to represent the plethora of plankton species existing in nature and has to organize those species into a few functional types represented by the model. However, such simplifications inherently leave out the potentially large variability of behaviours of the diverse species within each of the functional types. If the functional type internal variability (e.g. in sinking rates, metabolic rates, light absorption, trophic web-links) is large, the model

Table 2
The ERSEM parameters perturbed in this study. For further details on the parameter sensitivity analysis results see Stefano et al. (2022) and for the detailed overview of ERSEM parameters see Butenschön et al. (2016). Some abbreviations used are: dissolved organic matter - DOM, dissolved organic carbon - DOC, particulate organic matter - POM.

Related variable	Parameter	Original value from (Butenschön et al., 2016)
Bacteria	The efficiency at high oxygen levels	0.6
	The maximum turn-over rate of DOM	1 d^{-1}
	Fraction of semi-refractory DOC available to bacteria	0.0075
Diatoms	Threshold for nitrogen limitation (relative to Redfield ratio)	1
	Maximum specific productivity at reference temperature	1.375 d^{-1}
	Maximum nitrogen-to-carbon ratio	1.075
Dinoflagellates	Threshold for nitrogen limitation (relative to Redfield ratio)	1.0
Nanophytoplankton	Threshold for nitrogen limitation (relative to Redfield ratio)	1.0
Small-size POM	Sinking velocity	1 m/d

approximation of this diversity by a single, averaged functional type behaviour may be poor. This functional type behaviour is largely captured by specific model parameters, and the spatio-temporal variations of the model parameters can account for the spatio-temporal variations within internal functional type species composition, not resolved by the model. If the exact spatio-temporal parameter variation is unknown, one can still represent it stochastically, through spatio-temporally varying perturbations of the model parameters. The model then produces stochastic predictions aiming to accurately reproduce the distributions of the observed specific functional type behaviours. In this sense the spatio-temporally varying parameter perturbations can be considered to represent uncertainties of aleatory type.

In this work, we introduce into the NWES NEMO model spatio-temporally varying physics parameter perturbations using the stochastically perturbed parameters (SPP) scheme from Storto and Andriopoulos (2021), Lea et al. (2022). The SPP implementation is identical to the global implementation of Lea et al. (2022), to which the reader is directed for more details, except for the SPP perturbations relating to vertical mixing. In Storto and Andriopoulos (2021), Lea et al. (2022) perturbations were made to the turbulent kinetic energy (TKE) vertical mixing scheme within NEMO, but the NWES model uses an alternative vertical mixing scheme, generic length scale (GLS). Instead of the TKE perturbations (bottom three rows of Table 2 in Lea et al., 2022), the following GLS parameters were perturbed: GLS Galperin limit (rn_clim_galp), GLS EPS minimum (rn_epsmin), and GLS E minimum (rn_emin), all with a relative standard deviation of 0.1.

A future development could be to introduce similar spatio-temporally varying perturbations for the marine biogeochemistry model, once a careful analysis and validation of appropriate parameters and stochastic schemes (e.g. correlation length-scales for parameter perturbations) is possible.

There is another type of parameter uncertainty: imagine that even if we were modelling a system that could be perfectly described using spatio-temporally constant parameter values, we might easily lack precise knowledge about those constant parameter values. This lack of knowledge represents epistemic uncertainty and can be ideally reduced through parameter estimation tools, such as joint state-parameter estimation by means of data assimilation (e.g. Doron et al., 2011, 2013; Schartau et al., 2017; Gharamti et al., 2017a,b). In this work we developed such epistemic uncertainty representation of the (spatially and temporally constant) ERSEM model parameters. Since > 400 parameter uncertainties cannot be represented by O(10) member ensembles that we can computationally afford, we have firstly run a 1D analysis at two differing locations (L4 and BATS stations, Stefano et al., 2022) to select a set of nine most relevant parameters from the point of view of the selected seven MEHIs. These parameters were chosen based on their averaged sensitivity coefficient values across all the MEHIs (Saltelli et al., 2000, 2008), whilst making sure that, for each MEHI, the parameter to which the MEHI was most sensitive was included in the list. This is to minimize the risk that the uncertainties of certain MEHIs might be significantly underestimated relative to other MEHIs, due to a skewed choice of the perturbed parameters. The selected nine uncertain ERSEM parameters are shown in Table 2.

The parameter prior uncertainties were estimated by simultaneously perturbing all the parameters, i.e by drawing their values independently (a common practice, Schartau et al., 2017) from uniform distributions within a ±30% interval around their established values (see Table 2, also Butenschön et al., 2016). The prior PDFs for the parameters were based on an expert "first guess", that has been evaluated (leaving open the possibility of re-tuning the PDFs) through the ensemble validation described in Section 3.1.

2.2.3. Model structural uncertainties

The model structural uncertainties are related to the formulation of the model, such as limited spatial/temporal model resolution, misrepresented processes, or processes and variables that have not been included, or resolved in the model (e.g. Payne et al., 2016). Structural uncertainties and biases are perhaps the most significant sources of model uncertainty, especially in marine biogeochemistry/ecology, where representing the system complexity and diversity poses a great challenge, and many issues with the present generation of biogeochemical models have been pointed out (e.g. see Flynn et al., 2022). However, it is well known that it is very hard to estimate the structural model uncertainties, and they are typically considered among the "unknown unknowns" (e.g. Garrett, 1992).

In this work we partly addressed some structural uncertainties by implementing the NEMO model stochastic kinetic energy backscatter (SKEB) perturbations from the work of Storto and Andriopoulos (2021), Lea et al. (2022), representing the uncertainty in numerical and convective energy dissipation due to limited model spatial resolution.

Other than that, the most common way of trying to represent model structural uncertainties is running multi-model ensembles. Albeit computationally expensive and non-trivial to design (Payne et al., 2016), such ensembles are frequently considered in climate predictions

(e.g. O'Neill et al., 2016). One limitation of the multi-model ensemble approach is that some uncertainties cannot be well represented simply due to constraints on the computational power, e.g. all models from the ensemble will have limited spatial resolution, or have limited capability to represent biological complexity. Running multi-model ensembles is beyond the scope of this work and we therefore do not attempt to represent most of the structural model uncertainties here.

2.3. The data assimilation systems

In this section we introduce the existing deterministic variational operational system used at the Met Office, as well as a new development based on the ensemble representation of the uncertainties described in the previous section, the ensemble-variational system.

2.3.1. 3D variational system (3DVar)

NEMOVAR in its original version is a variational DA system (Mogensen et al., 2009, 2012; Waters et al., 2015), with its 3DVar version used in producing operational forecasts and reanalyses for the NWES. In marine biogeochemistry, the assimilation of ocean color (OC)-derived chlorophyll using NEMOVAR is highly successful in improving the NWES phytoplankton phenology, PFT community structure (using satellite PFT chlorophyll assimilation), underwater irradiance and to a more limited degree also carbon cycle representation (Skákala et al., 2018, 2020; Kay et al., 2019). NEMOVAR includes capability to assimilate multi-platform (satellite, in situ) data both for physics (e.g. Waters et al., 2015; King et al., 2018) and biogeochemistry (Ford, 2021; Skákala et al., 2021; Ford et al., 2022).

First Guess at Appropriate Time (FGAT) is used to calculate the innovations between the observed values and model background at the nearest model timestep to the observation times, during a 24 h forecast. Then NEMOVAR is used to produce a set of increments to update the model state variables. The increments are added into the model gradually over the same 24 h to avoid generating sudden shocks, using incremental analysis updates (IAU, Waters et al., 2015; King et al., 2018). In the physical DA application, NEMOVAR applies balancing relationships within the assimilation step and delivers a set of increments for temperature, salinity, sea surface height (SSH) and the horizontal velocity components. In its biogeochemical application it calculates a set of increments separately for each assimilated variable and subsequently balancing relationships are used to distribute those increments into a selected range of other ecosystem model variables. Within total chlorophyll assimilation, NEMOVAR first assimilates log-chlorophyll, since chlorophyll is considered to be log-normally distributed (Campbell, 1995). After the total chlorophyll increments are transformed from the log-space, the balancing scheme distributes the total chlorophyll increments into chlorophyll increments of the four PFTs, based on the simulated background community structure (Skákala et al., 2018), and the remaining PFT biomass components (carbon, phosphorus, nitrogen, silicon), based on the background stoichiometry.

NEMOVAR uses for both physics and biogeochemistry externally supplied, spatio-temporally varying observation error variances (the observation error correlations are unaccounted for), background error horizontal spatial covariances and background error variances. The background error horizontal spatial covariances are estimated as a weighted average of two Gaussian functions combining two horizontal correlation length-scales, one fixed at 100 km and the other based on the baroclinic Rossby radius of deformation (King et al., 2018). The spatially-varying weights between the two Gaussians are estimated separately for each variable. The background error variances originate from previous simulations and diagnostic assessments, e.g. for total chlorophyll the monthly climatologies are estimated based on a decadal Proudman Oceanographic Laboratory Coastal Ocean Modelling System (POLCOMS) - ERSEM ensemble run of Ciavatta et al. (2018). To inflate the background variances produced by the ensemble, the chlorophyll variances were re-calibrated to be 2–3 times larger than the observational error variances (Skákala et al., 2018). For biogeochemistry, both the background error variances and the horizontal spatial correlation length-scales were recently re-assessed and re-adjusted in Fowler et al. (2023) using statistics of Desroziers et al. (2005). These re-assessed background covariances from Fowler et al. (2023) are planned to be implemented in the future version of the Met Office operational system.

The vertical correlations are based on the scheme from Waters et al. (2015), King et al. (2018), Ford (2021), using flow-dependent vertical length scales, which are a linear function of depth until the base of the mixed layer and then scale with the spacing of the vertical layers in the model grid (for the details see the Eq.1 in Skákala et al., 2021).

2.3.2. 3D ensemble-variational system (3DEnVar)

This work uses the developments of Weaver et al. (2018), as implemented in the Met Office global physics system by Lea et al. (2022). The developments of Weaver et al. (2018) were implemented here for the coupled physical-biogeochemical model on the NWES domain. The features of the DA system described here remain mostly the same as in Lea et al. (2022). In this new 3DEnVar system, the ensemble based on perturbations from Section 2.2 is used, for both physics and biogeochemistry, to determine the background error variances and all the spatial covariances. In Lea et al. (2022) the background error variances and spatial covariances were calculated as weighted averages between the ensemble estimates and the parametrized climatology used in 3DVar. Here only the ensemble part is considered (the weight given to the ensemble-based background error covariances is 1), often referred to as "pure" 3DEnVar (Bannister, 2017; Lea et al., 2022). To avoid spurious horizontal spatial correlations, we use (as in Lea et al., 2022) localization lengthscales, that have been tuned for the NWES to 4°. The assimilation in the 3DEnVar system still uses the same balancing approaches as in the 3DVar system (as described in the previous section), ensemble estimates of cross-covariances between variables are not used.

Similarly to Lea et al. (2022), we included in the 3DEnVar system the perturbations to observations, which are required for the full consistency of the ensemble-based assimilation (e.g. Burgers et al., 1998; Isaksen et al., 2010; Bowler et al., 2017; Zuo et al., 2017; Van Leeuwen, 2020). Perturbations of both observational values and locations have been included. The observation values were perturbed similarly to Lea et al. (2022), using a white, Gaussian noise with the standard deviation set to a fixed value of 0.5 °C for temperature, 0.04 PSU for salinity, 0.6 $\log_{10}(\mathrm{m/m^3})$ for satellite log-chlorophyll and 0.3 $\log_{10}(\mathrm{mg/m^3})$ for glider log-chlorophyll. The observation locations were perturbed using a Gaussian distribution $N(0, 0.1°)$.

2.4. Assimilation experiments

In this work we have run two ensemble experiments called "free-bgc" run and "chl-DA" run. The only difference between the two simulations were the assimilated data listed in Table 3; i.e chl-DA assimilated physical data together with satellite and glider chlorophyll, while free-bgc assimilated only physical data. It is worth pointing out that the OC-CCI v5.0 chlorophyll data listed in Table 3 are provided with estimates of biases, and these were removed from the data before the assimilation using the method of Ciavatta et al. (2016). The chlorophyll glider data were from the AlterEco glider mission in the central North Sea (Loveday et al., 2022, see also Skákala et al., 2022) and the impact of glider chlorophyll profile assimilation largely remained localized in the glider mission area (e.g. see Skákala et al., 2022).

In all other aspects the two runs were identical: they were based on 30-member ensembles and used 3DEnVar for assimilation. The simulations were run for the biologically productive, March - September 2018 part of the year and initialized from the initial value conditions produced by a 2-month (January-February 2018) spin-up of the chl-DA simulation. The spin-up originated from initial values for 01/01/2018,

Table 3
The table shows the assimilated data in the two experiments from this study. The combined satellite sea surface temperature (SST) data-set consists of grouping together the following satellite products: GCOM-W1/AMSR-2, NOAA/AVHRR, MetOp/AVHRR, MSG/SEVIRI, Sentinel-3/SLSTR and Suomi-NPP/VIIRS. The abbreviations are: "T": temperature, "S": salinity, "GTS": Global Telecommunication System, "OC-CCI": Ocean Color - Climate Change Initiative of the European Space Agency (ESA), (for the details see Sathyendranath et al. (2019)).

Experiment	Physics satellite	Physics in situ	Biogeochemistry satellite	Biogeochemistry in situ
free-bgc	Combined SST data-set	EN4 data-set for T& S SST from GTS	–	–
chl-DA	Combined SST data-set	EN4 data for T& S SST from GTS	OC-CCI v5.0 chlorophyll	Glider chlorophyll

which were identical to all the ensemble members and obtained from a coupled physical-biogeochemical reanalysis produced in Skákala et al. (2022). The different members in the 30-member marine model ensemble were distinguished by different NEMO and ERSEM parameter values, as well as initial value conditions, following the spin-up. They were also typically forced by different members of the 10-member ERA5 atmospheric ensemble, with each ERA5 member acting as forcing for three marine model members (as the size of ERA5 ensemble was exactly 1/3 of the size of the marine model ensemble).

2.5. Validation tools, data and diagnostics

2.5.1. The reliability score for ensemble design

The aim of the ensemble is to represent the uncertainty in the data assimilation system and its forecast. The method of Rodwell et al. (2016) validates the reliability of the ensemble estimate of the uncertainty in the background (the 1-day forecast) by partitioning the variance of the background-observation misfits into its different components. The background-observation misfit, commonly referred to as the innovation ($\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x_b}$, with $\mathbf{y}$ denoting the observations, $\mathbf{x_b}$ the background and $\mathbf{H}$ the observation operator), can be also written as the difference between the error in the observations $\epsilon_y \sim N(\beta_y, \mathbf{R})$ and the error in the background $\epsilon_b \sim N(\beta_b, \mathbf{B})$. Therefore, the variance of the innovation is given by

$$Var(\mathbf{d}) = \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}} + \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathbf{T}}, \tag{1}$$

where $\boldsymbol{\beta} = \beta_y - \beta_b$ is the difference in the observation and background biases or the mean innovation, $\mathbf{B}$ is the background error covariance matrix and $\mathbf{R}$ is the observation error matrix.

In the ensemble DA system developed, $\mathbf{B}$ is approximated by the sample estimate of the variance of the ensemble. Therefore the reliability of the ensemble can be assessed by quantifying the difference (residual) between the left and right-hand sides of Eq. (1). The residual is defined such that a negative value implies that the ensemble overestimates the uncertainty, whereas a positive value implies that the ensemble underestimates the uncertainty. This interpretation of the residual statistics is slightly clouded by the fact that $\mathbf{R}$, the observation error variance, is also not known exactly. Within this work $\mathbf{R}$ in Eq. (1) is approximated as the sample estimate of the variance of the perturbed observations, which is slightly different to the error variance assumed during the assimilation.

Lea et al. (2022) used this method to validate the ensemble DA system for the physical component of the global ocean system using observations of sea level anomalies, sea surface temperature, potential temperature and salinity. Here we apply this to sea surface chlorophyll observations to validate the coupled biogeochemistry component on the NWES.

In addition to the metrics from Rodwell et al. (2016), we used rank histograms (e.g. Hamill, 2001) to evaluate the physical part of the model ensemble. Rank histograms are based on comparing cumulative distribution functions (CDFs) between the observations and the ensemble forecast. The histogram shows numbers of observed values that fall into the intervals between the ordered ensemble forecast values. It is evaluated at each time and spatial location, and then aggregated across all times and spatial locations. A flat rank histogram might indicate a good match between the observed and the ensemble CDFs. Rank histograms are a straightforward way of evaluating aleatory-type uncertainties, which dominate the physical model perturbations used here (see the discussion in Section 2.2).

2.5.2. Observability metrics

By the observability of MEHI by total chlorophyll we broadly mean the capacity of chlorophyll observations to constrain the simulations of MEHIs. Rather than exploring the impact of a particular observation at a specific location and time, we are interested in an aggregate impact of total chlorophyll observations on MEHI dynamics across the whole NWES and a longer period of time.

There exist many theoretical results on observability (e.g. Moore, 1981; Kang and Xu, 2009, also in relation to DA (King et al., 2015, 2017). Although these have now been extended to time-varying non-linear systems (Martinelli, 2022) and models with larger ($O(10)$) numbers of degrees of freedom (Stigter et al., 2017), it remains essentially impossible to apply these tools for the MEHI analysis from this study, which connects many cycles of DA with a model having $O(10^7)$ degrees of freedom.

We will proceed with defining an observability metric that can be more easily applied to the aggregated outputs of the high-dimensional model used here. To do this, we will start with the concept of ignorance score from Roulston and Smith (2002): Let us consider a PDF p corresponding to a target distribution of MEHI values observed in nature. Then if q is a PDF by which we try to estimate the target distribution p, the information cross-entropy $H(p, q)$ is a good measure for how well q estimates p:

$$H(p, q) = -\sum_i p_i \log_2(q_i). \tag{2}$$

The score from Eq. (2) has minimum when $p = q$, measuring the information entropy of the target PDF p (the minimum number of bits needed to code the information in p), whilst the deviation of $H(p, q)$ from this minimum gives the redundant bits of information due to imperfect representation of p by q. In Germineaud et al. (2019) a normalized (to the [0,1] range) ratio of $H(p, p)/H(p, q)$ was introduced, which attains zero value in case of maximum distance between p and q, whilst reaching the maximum value (=1) for a perfect estimate of p by q ($q = p$).

The general experience (see Skákala et al., 2022, but also results from this work) is that there is a significant mismatch between the simulated NWES chlorophyll in the ERSEM free run and the available observations. Therefore assimilating chlorophyll observations into the model leads to major changes to the simulated chlorophyll values. One could then expect that, unless there is a major error cancellation, a strong link between MEHI and chlorophyll (required by the observability of the MEHI through chlorophyll) would imply that the MEHI free run PDF q would also be significantly mismatched with the analysis MEHI PDF p. If this is true, then the MEHI's observability through the assimilated OC chlorophyll data will be proportional to the change in the MEHI's PDF as a result of assimilating OC chlorophyll into the

model. The observability score $\mathrm{Obs}(v,x,t)$ of MEHI v at spatial location x and time t we will then define as

$$\mathrm{Obs}(v,x,t) = 1 - \frac{H(p(v,x,t),p(v,x,t))}{H(q(v,x,t),p(v,x,t))}, \tag{3}$$

where $p(v,x,t)$ is the target PDF of v represented by the analysis PDF (chl-DA run), whilst $q(v,x,t)$ is the free run PDF of v (free-bgc run), (both at the spatial location x and time t).

The observability score from Eq. (3) in fact measures the distance between $p(v,x,t)$ and $q(v,x,t)$ known as a relative entropy, or Kullback–Leibler divergence measure (Kullback and Leibler, 1951), normalized by cross-entropy $H(p,q)$. $\mathrm{Obs}(v,x,t)$ metric is also constrained to the interval of [0,1] values: When the match between the free run and the OC chlorophyll analysis PDFs (for v at x,t) is perfect, the $\mathrm{Obs}(v,x,t)$ score gives zero value, which means assimilating observations of OC chlorophyll did not manage to modify the MEHI v's PDF. When the mismatch between the analysis and the free run PDF is maximum, then MEHI v is perfectly observable and $\mathrm{Obs}(v,x,t) = 1$.

To support the conclusions about observability, that are based on metrics from Eq. (3), we compare the results with an independent metric for observability, defined as

$$\mathrm{Obs}^{(2)}(v,x,t) = \frac{\mathrm{RMSD}(v,x,t)}{\sigma_e(v,x,t)}. \tag{4}$$

Here $\mathrm{RMSD}(v,x,t)$ is the Root Mean Squared Difference calculated for MEHI v at the location x and time t, by (i) comparing the differences between the corresponding ensemble members of the chl-DA run and the free-bgc run and (ii) averaging them across the ensemble. The $\sigma_e(v,x,t)$ is the standard deviation of MEHI v in the free-bgc run ensemble at the same location and time. The metric $\mathrm{Obs}^{(2)}(v,x,t)$ can be interpreted as measuring the averaged change to MEHI v due to re-initializing the model with assimilated total chlorophyll observations, relative to how sensitive v is to the other model perturbations.

2.5.3. Uncertainty measure

The uncertainty measure of MEHI $\mathrm{Un}(v,x,t)$ we define as

$$\mathrm{Un}(v,x,t) = \frac{\sigma_e^2(v,x,t)}{\sigma_e^2(v,x,t) + \sigma_s^2(v,t)}, \tag{5}$$

where σ_e is the standard deviation of the ensemble of MEHI v at the spatial location x and time t. σ_s is the standard deviation (at time t) calculated across the NWES domain for the ensemble mean of MEHI v (describes spatial variability of the MEHI ensemble mean).

The definition from Eq. (5) is motivated by two considerations: (i) the need to normalize uncertainty, similarly to the observability, to the [0,1] interval, so that these two measures have the same range of values and can be reasonably combined to estimate the need for observations (see Section 2.5.4). (ii) $\mathrm{Un}(v,x,t)$ is naturally interpreted as the proportion of uncertainty-related variability within the total MEHI v variability at time t. The total variability of the ensemble is due to both spatial variability of the ensemble mean values and the internal variability (or spread) of the ensemble representing the model uncertainty. Another way of looking at the measure in Eq. (5) is that it gives the noise-to-signal ratio within the spatial maps of MEHIs, which are provided every day by an operational system as an analysis, or forecast.

The Eq. (5) has a straightforward interpretation for normally distributed variables: If we had no prior knowledge of how MEHI v is distributed on the NWES at time t, but wanted to statistically predict its value in some specific location x from PDF f, then if both the histograms of ensemble mean values (across NWES) and the ensemble distribution at x were Gaussian, the $\mathrm{Un}(v,x,t)$ metric would just give the ratio of the variance of the uncertainty (error) component at x to the variance of f.

Although MEHI uncertainty can be significant on the model resolution scale, it is generally expected that for the coarser scale MEHI values (e.g. NWES-wide mean values) the uncertainty is significantly reduced. This is why it is easier to predict global phenomena, such as global mean temperatures over climatological scales, than local weather patterns. We will investigate here how "localized" the MEHI uncertainty is, i.e. how much it reduces for MEHI NWES-wide average values compared to the MEHI uncertainty on the 7 km model spatial resolution scale. To do that we use a very similar metric to Eq. (5):

$$\mathrm{Sc}(v,t) = \frac{\sigma_{av}^2(v,t)}{\sigma_{av}^2(v,t) + \sigma_{fl}^2(v,t)}. \tag{6}$$

In Eq. (6) we decomposed the total uncertainty of MEHI v into the uncertainty of the NWES-wide average value (at time t), $\sigma_{av}(v,t)$, and the uncertainty associated with fluctuations on the model grid spatial resolution scale, $\sigma_{fl}(v,t)$. The $\sigma_{av}(v,t)$ is calculated from the standard deviation of NWES-wide mean values of each ensemble member (at time t):

$$\sigma_{av}(v,t) = \sigma_e(\langle v_i(x,t)\rangle_x), \tag{7}$$

where v_i is an ensemble member of MEHI v (indexed by i) and $\langle\rangle_x$ means spatial averaging through the NWES. Furthermore, σ_e is the standard deviation across the ensemble, as defined in Eq. (5). The $\sigma_{fl}(v,t)$ was computed by subtracting from each ensemble member v_i its NWES-wide difference between the v_i and the ensemble mean:

$$B_i(t) = \langle v_i(x,t)\rangle_x - \langle\langle v_j(x,t)\rangle_x\rangle_j, \tag{8}$$

where $\langle\rangle_j$ means averaging through the ensemble. After the NWES-wide difference $B_i(t)$ was subtracted from each MEHI ensemble member v_i, $\sigma_{fl}(v,t)$ is calculated as an NWES average across the ensemble standard deviation of the MEHI corrected for $B_i(t)$, $v_i^{bc}, (v_i^{bc}(x,t) = v_i(x,t) - B_i(t))$:

$$\sigma_{fl}(v,t) = \langle\sigma_e(v_i^{bc}(x,t))\rangle_x. \tag{9}$$

The metric $\mathrm{Sc}(v,x,t)$ from Eq. (6) tells us about the proportion of NWES-wide (large scale) MEHI uncertainty within the total uncertainty. It can also be interpreted for two Gaussian distributions similarly to the metric from Eq. (5): Eq. (6) would give the ratio of the variance of the MEHI NWES-wide mean to the variance of the distribution predicting the MEHI at a specific spatial location.

2.5.4. The need for observations

Finally, the need for observations at location x and time t is directly proportional to the MEHI uncertainty and inversely proportional to its observability. Thus the metric for the "need for observations" $\mathrm{NO}(v,x,t)$ we propose to define here as a simple average between the uncertainty $\mathrm{Un}(v,x,t)$ (Eq. (5)) and the non-observability $1-\mathrm{Obs}(v,x,t)$ (see Eq. (3)):

$$\mathrm{NO}(v,x,t) = 0.5\cdot\mathrm{Un}(v,x,t) + 0.5\cdot[1-\mathrm{Obs}(v,x,t)]. \tag{10}$$

2.5.5. The validation data

Due to the assimilation of both satellite and in situ physics and chlorophyll observations, there were limited independent data left for validation. We have done validation using the observations at the L4 station, which is part of the Western Channel Observatory (WCO, https://www.westernchannelobservatory.org.uk/) and has one of the longest time-series of marine biogeochemical measurements in the world (Harris, 2010). The station is located in the English Channel, about 15 km south of Plymouth. It provides data typically with 5–7 day sampling frequency and we used the data for temperature, salinity, total chlorophyll, oxygen, nitrate, phosphate, silicate and ammonium. We have also used the data from the North Sea Biogeochemistry Climatology (NSBC) project (Hinrichs et al., 2017). The NSBC climatologies are based on in situ measurements collected across the NWES domain through the period of 1960–2014. The NSBC data have 1/4° spatial resolution.

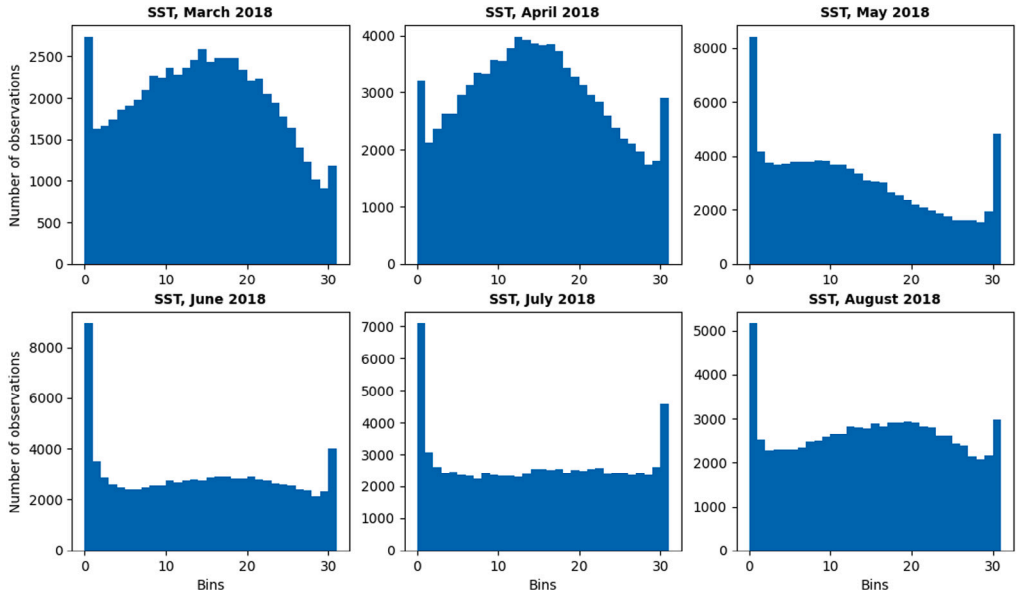

Fig. 1. Rank histogram for the SST comparing the analysis from the free-bgc run with the satellite observations across the NWES. To account for satellite observation SST errors, the model ensemble outputs were randomly perturbed by the observation errors (a standard technique to deal with this issue, Bowler, 2008).

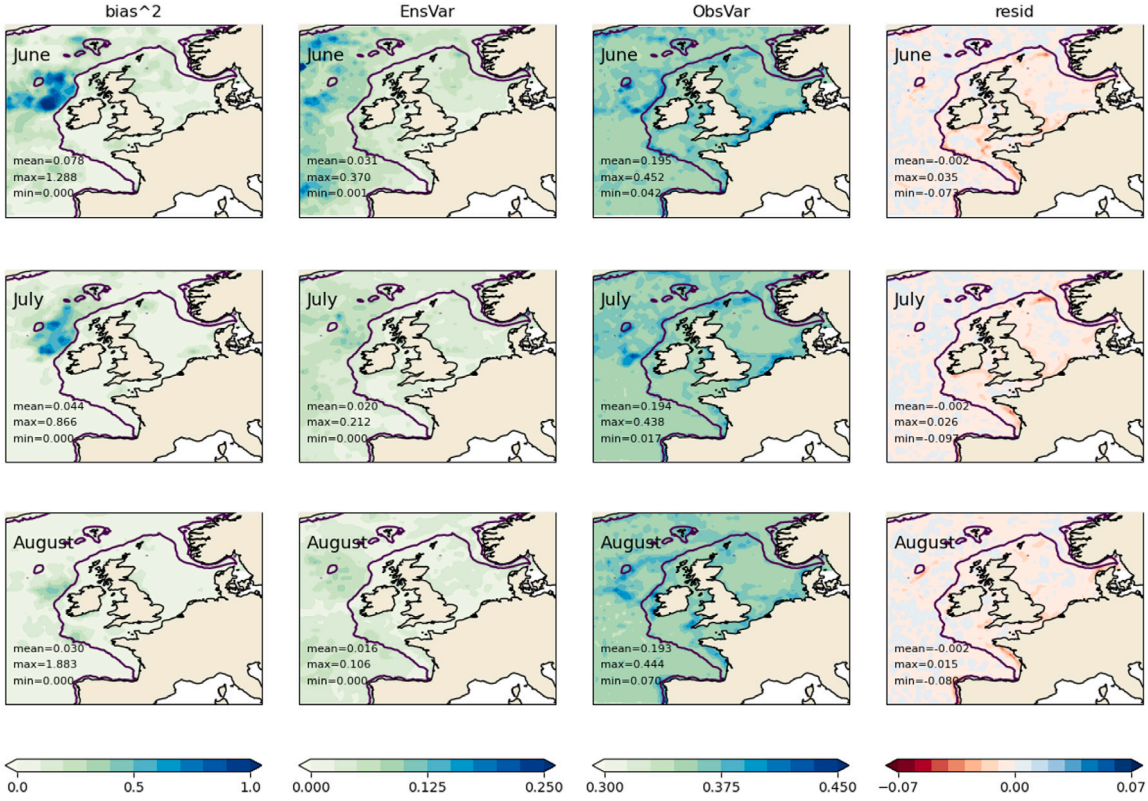

Fig. 2. Reliability statistics of the total surface chlorophyll-*a* from the ensemble validated against ocean colour data for June (top row), July (middle row) and August (bottom row). From left to right the variables plotted are the diagonal elements of $\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$, ensemble estimate of the background error variances, the sample estimate of the observation error variances from the perturbed observations, and the residual difference with the variance of the innovations (see Eq. (1)). The units are $\log_{10}^{2}(\mathrm{mg/m^3})$. The NWES region is delimited by the bathymetry being less than 200 m and marked by the black contour.

3. Results and discussion

3.1. Validation of the ensemble spread

The physical part of the ensemble was validated using a rank histogram (Hamill, 2001; Casati et al., 2008) comparing the SST from the analysis with the satellite observations. The rank histogram is shown in Fig. 1, demonstrating that the skill of the physical model ensemble spread exhibits significant spatio-temporal variation, i.e. the complexity of the histograms indicate that both the SST biases, and whether the physical ensemble is over-spread, or under-spread, strongly depend on the location and time. This suggests that the physical ensemble perturbations need improvement, but also that the improvement might

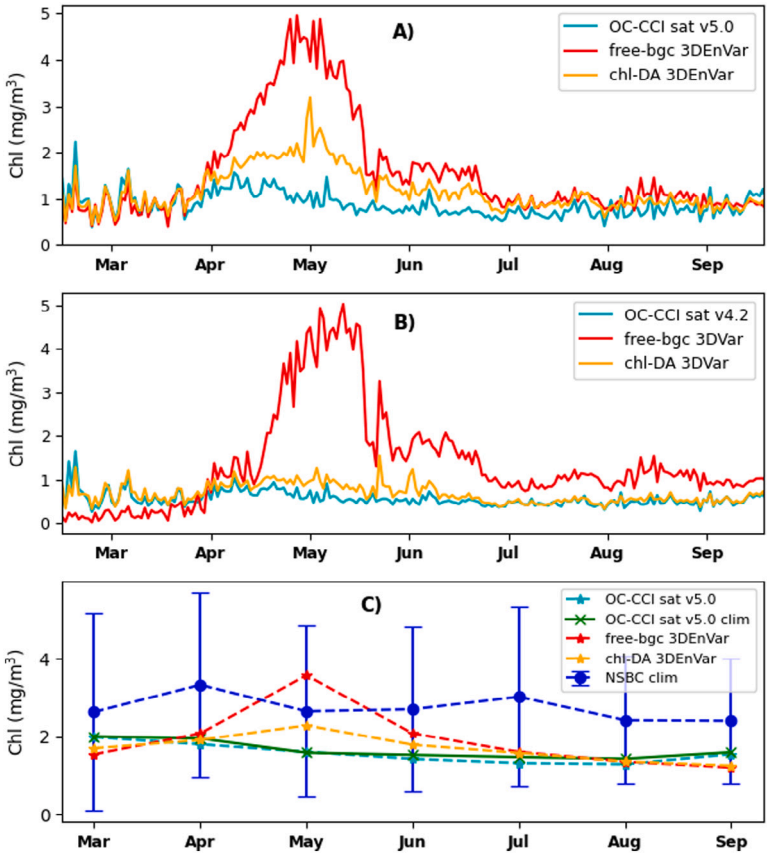

Fig. 3. Time series of surface total chlorophyll-*a* (mg/m^3) averaged across the NWES (bathymetry < 200 m). The upper panel (A) compares the ensemble means of the free-bgc run with the chl-DA run ("3DEnVar" is added to run names to distinguish that the assimilation was done with 3DEnVar), and the bias-corrected OC-CCI v5.0 chlorophyll data assimilated in the chl-DA 3DEnVar run. The middle panel (B) compares the analogues of free-bgc and chl-DA experiments based on the currently operational 3DVar system and the bias-corrected OC-CCI v4.2 chlorophyll data assimilated in the chl-DA 3DVar run (the 3DVar simulations are taken from Skákala et al. (2022) and the chl-DA assimilated an older, v4.2, OC-CCI chlorophyll product). The outputs of the model simulations were masked wherever there were missing satellite observations, ensuring the averages can be directly compared. The bottom panel (C) shows the NWES average of NSBC 1960–2014 surface chlorophyll monthly climatology, with the error bars corresponding to the NWES-average of the NSBC data standard deviation, describing the inter-annual variability in the surface chlorophyll data at the NSBC 1/4° resolution scale. The NSBC climatology is compared with the corresponding monthly averages of bias-corrected OC-CCI v5.0 satellite data and ensemble means of free-bgc and the chl-DA 3DEnVar, all taken at the NSBC data locations. To see how the 2018 chlorophyll might have deviated from its long term climatological values, we have also plotted the 2010–2019 climatology of the bias-corrected OC-CCI v5.0 chlorophyll.

need to be highly non-trivial, such as introducing spatial and temporal variability into the size of the NEMO model perturbations. This improvement will need to be addressed in the future, but appears to have only limited impact on the biogeochemistry model uncertainties, as shown by the validation of the chlorophyll uncertainty that follows (the MEHIs from this study have much more direct dynamical link to chlorophyll than to SST).

The reliability of the biogeochemistry component was validated using the method of Rodwell et al. (2016) described in Section 2.5.1. The results for the months of June to August are shown in Fig. 2. The residual values (last column) show that the ensemble spread for the observed surface chlorophyll is reliable for most of the time (small values compared to the error variances). The conclusion that the ensemble is reliable is supported by the good agreement between the background error variances estimated by the ensemble and the estimates of **B** computed using consistency diagnostics in Fowler et al. (2023), both in terms of magnitudes and spatial variability.

When structures are seen in the residual statistics these tend to be close to the edge of the shelf and are negative, indicating overspread of the estimated innovation variance. This should be noted when drawing conclusions about the need for observations in these regions and at these times. However, the regions of negative residual values correspond more closely to large estimates of the observation error variance rather than the background error variance. As such the negative residual values may be due to the way the observations are perturbed and so the ensemble of the model is still reliable for estimating the metrics of interest but could be improved by tuning the observation perturbations.

Large biases in the innovations are seen in North-West approach, an area in which the model is known to produce too large blooms in the late summer (Fowler et al., 2023). This bias, although significant, does not impact the metrics used to assess the observability and uncertainty of the MEHIs.

The reliability statistics were also estimated for the other months in the trial but are not shown because the values were always insignificant compared to the error variances, with no apparent structure.

3.2. Validation of the impact of ensemble-variational data assimilation

Estimates of MEHI uncertainty and observability are DA-system dependent and it makes little sense to evaluate them for a system that lacks skill. In this section we discuss some validation of the 3DEnVar system used in this study, and in some instances we compare it with the 3DVar system presently run operationally at the Met Office (the 3DVar data are taken from simulations of Skákala et al., 2022).

Firstly, the ensemble uncertainties fed into the background error covariance matrix in the 3DEnVar system were for the total chlorophyll on average 7–8 times lower than the parametrized uncertainties used in the 3DVar system (not shown here). This is broadly consistent with the diagnostics of Fowler et al. (2023) indicating that the parametrized

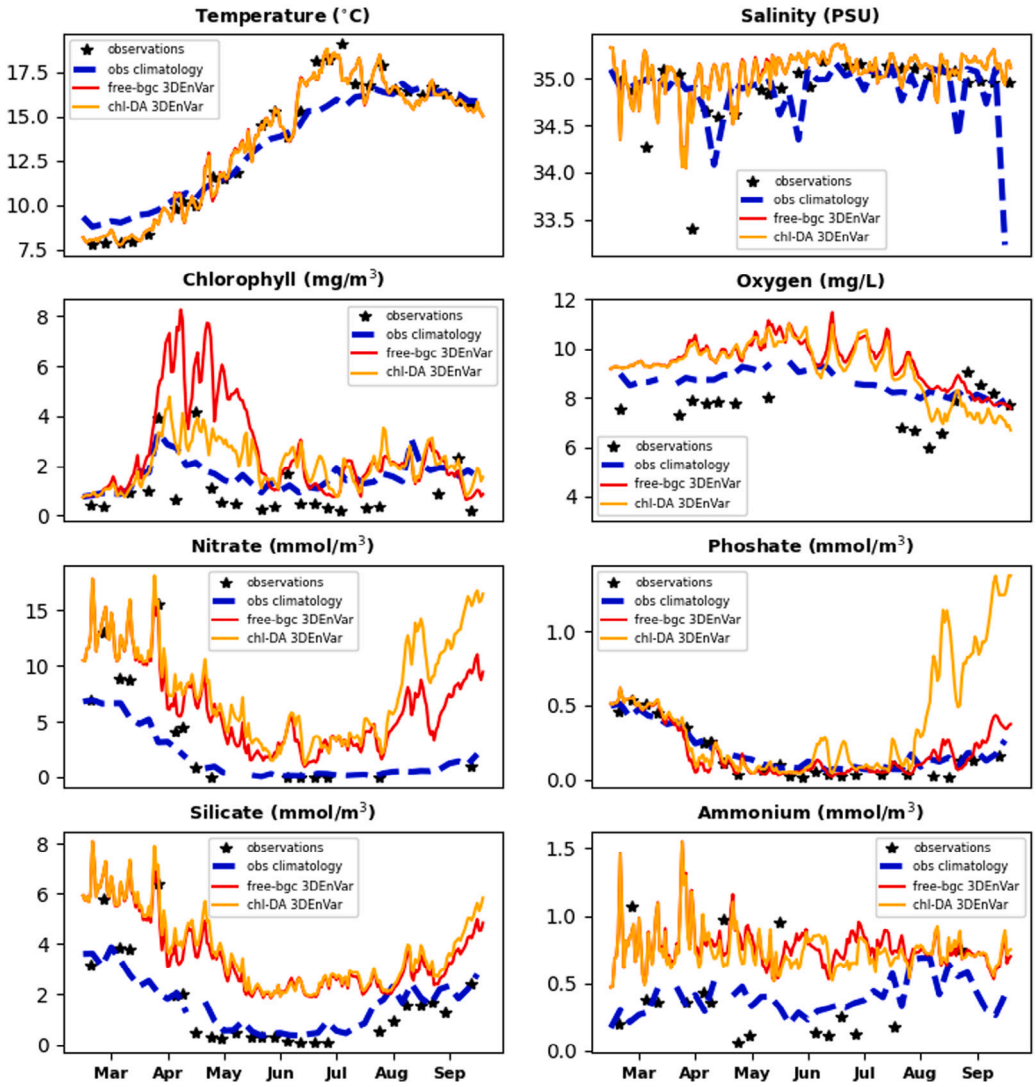

Fig. 4. Validation of the simulations at the L4 station, comparing the free-bgc run with the chl-DA 3DEnVar run (the same notation as in Fig. 3). The simulated surface values are compared with the corresponding L4 surface observations and with the L4 surface observational climatology.

total chlorophyll background error variances used by the currently operational 3DVar Met Office system are substantial overestimates.

The 3DVar system is known to produce reanalyses that converge very closely to the assimilated observations, when compared to the free run (Skákala et al., 2018, 2020, 2021, 2022). This can be seen also in the panel B of Fig. 3. The proximity of the analysis to the assimilated observations is not necessarily optimal (Skákala et al., 2020, 2021, 2022): although the assimilated (mostly satellite) observations are considered much more trustworthy than the model free run (Skákala et al., 2018, 2022), they are known to significantly deviate from the "true state" (e.g. Skákala et al., 2021; Ford et al., 2022), which is typically thought to be better represented by the in situ observations. It has been suggested in Skákala et al. (2022), that the close match between the reanalysis and the assimilated observations (relative to the free run) is a combined consequence of (i) the lack of ability to capture the reduction of background error as the analysis approaches the true state (due to lack of flow-dependence in the system) and (ii) the not entirely correct (e.g. see Fowler et al., 2023) assumption that observation errors are uncorrelated, artificially increasing the impact of observations. Comparison between Fig. 3(A) and Fig. 3(B) indeed shows that using flow-dependent background error variances in the 3DEnVar system increases the gap between the analysis state and the assimilated observations relative to 3DVar (plots that include the vertical dimension are shown in Fig.S1 of the Supporting Information (SI)). At the same time the validation of the ensemble spread (Fig. 2) indicates that this is likely not due to the ensemble losing its spread and the model becoming overconfident, a common issue in ensemble methods (e.g. Slingo and Palmer, 2011). The panel C of Fig. 3 compares satellite surface chlorophyll to NSBC climatology and indicates that the assimilated satellite data might be underestimating chlorophyll concentrations in the NSBC locations, suggesting that the deviation of the analysis from the assimilated satellite data in the late Spring (Fig. 3:C) could perhaps move the analysis closer to the true state. The use of NSBC monthly climatology might be seen as problematic, due to high inter-annual variability of the NSBC data (Fig. 3:C), however comparing the 2018 satellite data to satellite climatology in Fig. 3:C suggests that the 2018 year might have been reasonably representative of the longer-term climatological data.

Fig. 4 presents validation of the 3DEnVar system using the observations and their seasonal climatology from the specific location at the L4 station. The climatology is particularly relevant for chlorophyll, which can be highly variable relative to the L4 sampling frequency, and oxygen where the 2018 measurements seemed to be highly anomalous and thus potentially of questionable quality (for both see Skakala et al., 2023). Interestingly, the chlorophyll L4 observational climatology matches very nicely with the chl-DA run produced by the 3DEnVar

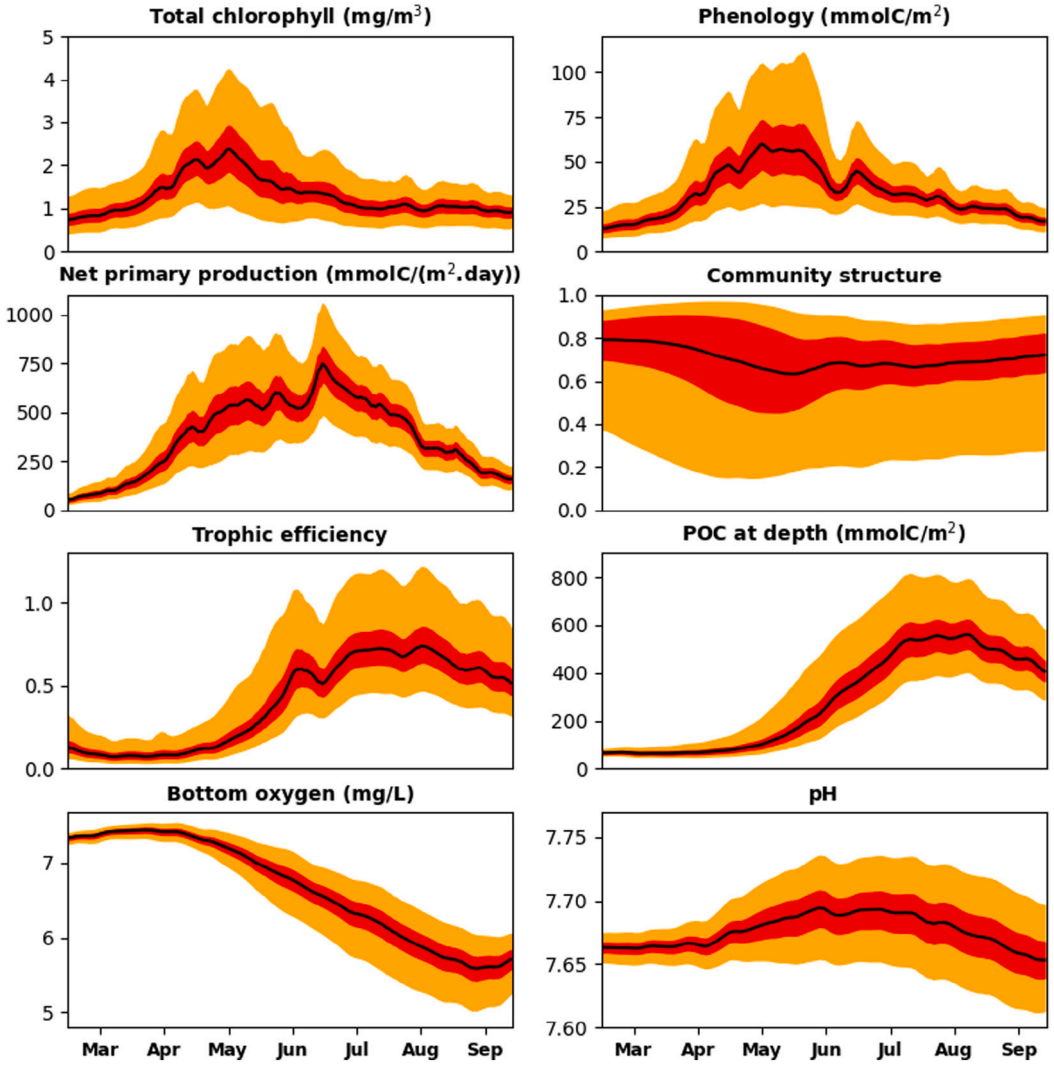

Fig. 5. The uncertainties of MEHIs in the chl-DA experiment using 3DEnVar. The time series are for the averaged values across the NWES (bathymetry < 200 m), the black line always represents the ensemble median, the red colour shows the two quartiles around the median value (25% of members beneath and above the median) and the orange colour shows the full ensemble spread.

system. The phytoplankton spring bloom simulated by 3DEnVar reproduces nicely the 2018 L4 observations, but both free-bgc and chl-DA 3DEnVar simulations overestimate chlorophyll for the rest of the year. The results of the nutrients validation are quite mixed, with some signs that the 3DEnVar chl-DA run might overestimate significantly phosphate (and to a lesser degree also nitrate) towards the end of the simulation period. The increased phosphate values at the end of the simulation period appeared in larger parts of the domain and across all ensemble members, however the cause could not be identified from the existing model outputs. The increase of phosphate in September has no impact on the results presented in this study and will need to be investigated later with more detailed outputs and validation tools.

3.3. Marine ecosystem health indicators' (MEHIs) uncertainties and observability

Fig. 5 shows time series of the ensemble distribution (described in quartiles) of the spatial mean across the NWES for each MEHI. The spatial distributions of the MEHI ensemble means and standard deviations are shown in Figs. 6 and 7. Figs. 6–7 demonstrate that the MEHIs have typically much lower uncertainties on the NWES than in the open ocean region. This is welcome news for the product users, as NWES biogeochemistry is mostly of greater end-user interest than in the open ocean. The lower MEHI uncertainty on the shelf can be seen also in the relative uncertainties (the absolute uncertainties from Figs. 6–7 normalized by the mean values), shown in Fig.S2 of SI. In particular, with the exception of oxygen and pH, all the MEHIs had at least two times lower relative uncertainty on the NWES than in the open ocean (Fig.S2 of SI). The largest difference in uncertainty between the NWES and the open ocean was for trophic efficiency and net primary production, which had uncertainty more than an order of magnitude lower on the NWES than in the open ocean (Fig. 6).

All the MEHIs except for oxygen and pH have on the NWES relative uncertainty between 20%–40% (Fig.S2 of SI), with the relative uncertainty of oxygen and pH somewhat lower (oxygen $\sim$ 8% and pH $< 1\%$). However, more relevant is to compare the MEHI uncertainties defined in Eq. (5) ($\mathrm{Un}(v,x,t)$). The results are shown in Fig. 8(A), demonstrating that with the exception of phytoplankton community structure, for all MEHI's the uncertainty is a relatively small (always $< 20\%$, and mostly $< 10\%$) part of total variability (see Eq. (5) and the discussion therein). This suggests that spatial maps of MEHIs derived from the daily analyses by the Met Office (with some extrapolation this

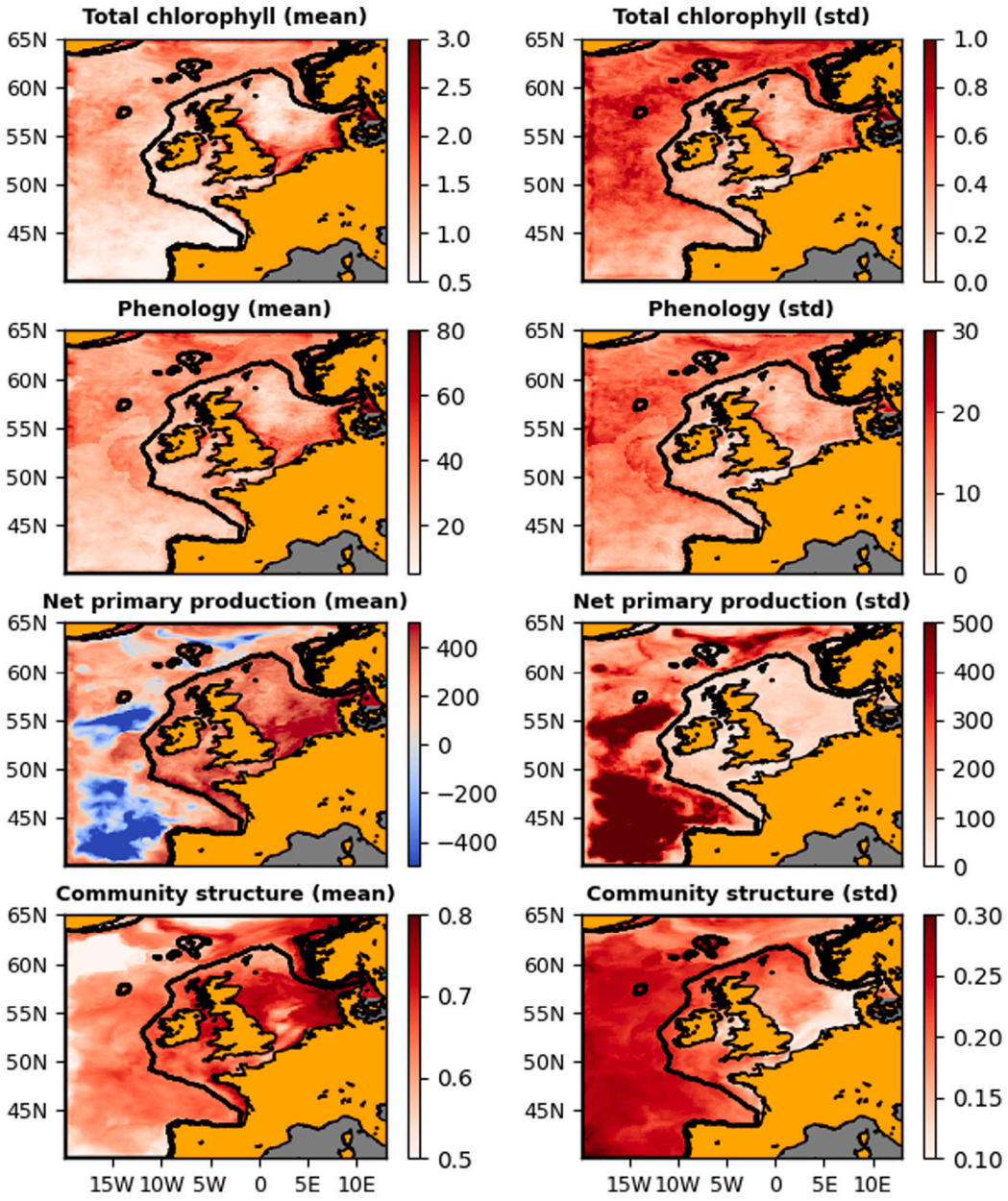

Fig. 6. The mean values (left-hand panels) and the uncertainties measured by the ensemble standard deviation (right-hand panels) of four MEHIs from the chl-DA experiment. The maps show averages across the March-September 2018 period. The units are for each MEHI the same as in Fig. 5. Although in some of the plots one can see the imprint of the open lateral boundary conditions on the MEHI ensemble mean and uncertainty, this effect remains localized far from the NWES.

should be also true about short-range forecasts), can be mostly trusted. The key exception is the phytoplankton community structure, where uncertainty dominates spatial variability, suggesting that the products for phytoplankton community structure derived from the assimilation of total chlorophyll are fundamentally unreliable. However, it should be noted that satellite OC-derived products of chlorophyll partitioned in the 4 ERSEM PFT groups already exist (Brewin et al., 2017), as well as the capability to assimilate those products within the Met Office system (Skákala et al., 2018). This work therefore highlights the importance of those satellite PFT chlorophyll products and their assimilation into the NWES marine biogeochemistry model.

Fig. 8(B) shows the fraction of the total uncertainty that is retained on the NWES-wide scale (metric from Eq. (6) and the discussion therein). As discussed in Section 2.5.3, the wider-domain averages of MEHIs will be typically more certain than their values on finer resolution scale. Fig. 8(B) then shows that by taking NWES-wide MEHI averages rather than their values on the 7 km product grid-scale, one reduces the MEHI uncertainty by 70%–95%. It is also demonstrated that from the seven MEHIs, the phytoplankton community structure uncertainty is the most "non-local", i.e. 30% of its uncertainty remains on the NWES-wide scale. In Fig.S3 of SI we show how the uncertainty is further reduced if we average the MEHIs not just in space, but also in time. The MEHI uncertainty is reduced through time-averaging across the simulation period by another 50%–60%, with only small differences among the MEHIs (Fig.S3 of SI).

The need for phytoplankton community structure observations is further emphasized by Fig. 8:C comparing the observability (Eq. (3)) of the different MEHIs from the OC-derived total chlorophyll surface

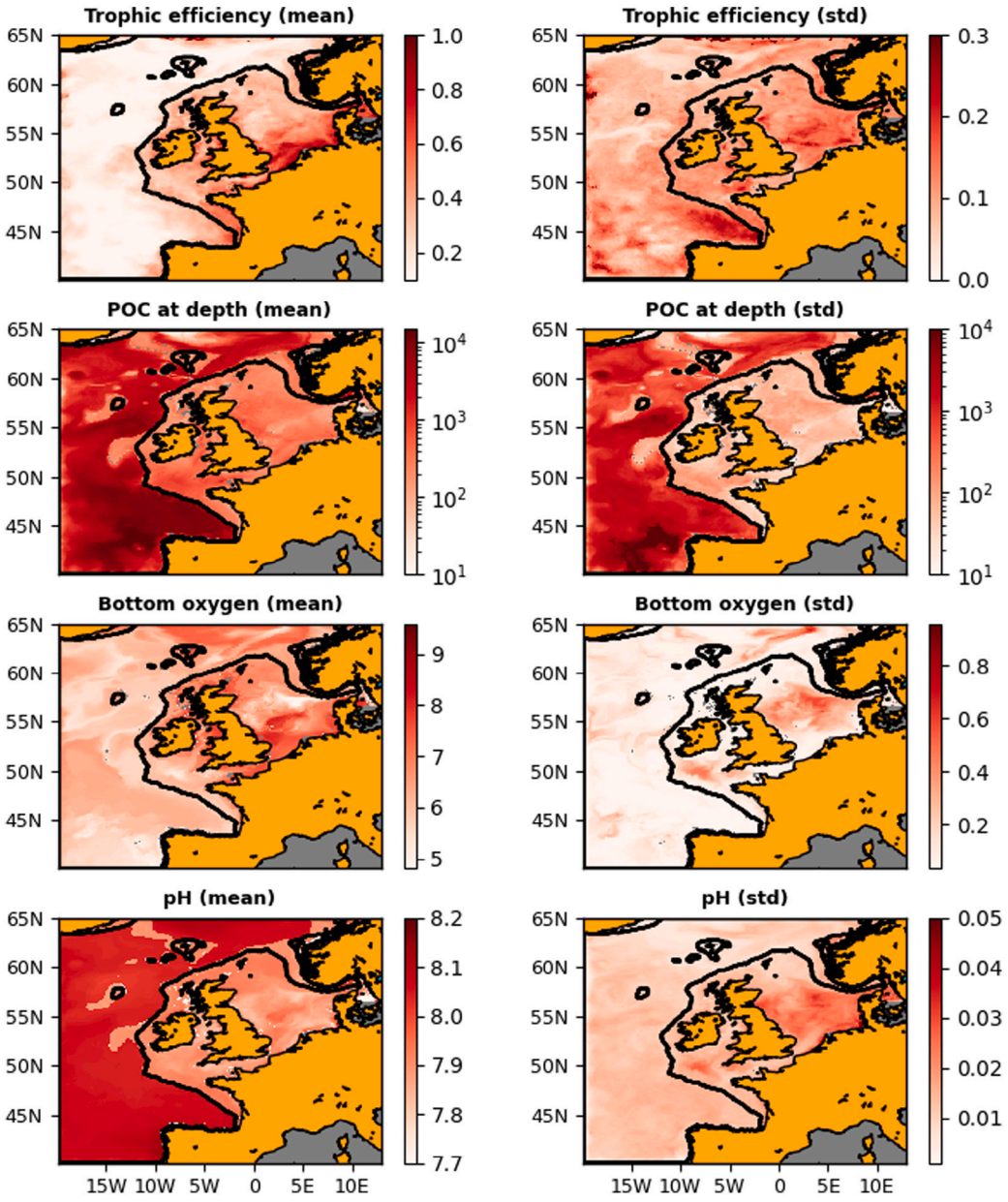

Fig. 7. The same as in Fig. 6, but for the remaining four MEHI's. The MEHI units are the same as in Fig. 5.

concentrations. Not surprisingly, the most observable is phytoplankton phenology together with the trophic efficiency (which is closely linked to phenology), but other MEHIs are lagging not very far behind. The least observable are the pH and the phytoplankton community structure, which is also confirmed by the alternative observability metric $\mathrm{Obs}^{(2)}(v,x,t)$ from Eq. (4), shown in Fig.S4 of SI. The low observability and the high uncertainty of the phytoplankton community structure lead to a high $\mathrm{NO}(v,x,t)$ score shown in Fig. 8:C.

Fig. 9 shows the spatial locations on the NWES where MEHI observations are needed, based on the spatial distribution of $\mathrm{NO}(v,x,t)$ score (Eq. (10)). The areas where the need was highest are highlighted with cyan colour. These areas are in most cases in the coastal zone, typically around the UK and Ireland. The high need of observations in the coastal zones ($\mathrm{NO}(v,x,t)$, Fig. 9) is due to low observability score (not shown here, but it can be deduced from $\mathrm{NO}(v,x,t)$ and the uncertainties in Figs. 6–7). The low observability near the coast might be caused by the fact that biogeochemistry in the coastal regions highly depends on the riverine discharge, but rivers might impact differently the MEHI's and the observed phytoplankton chlorophyll. For example, changes in sea salinity caused by the freshwater river flux can have a large impact on pH, but might not affect chlorophyll in the same way. Furthermore, the satellite data uncertainties are typically larger in the coastal zone, which reduces the overall impact of chlorophyll assimilation, hence MEHI observability.

4. Conclusions

This work provided uncertainty estimates of a selected class of marine ecosystem health indicators (MEHIs) on the North-West European Shelf (NWES). This is of major significance: the uncertainty information is essential for the end-users and the data assimilation, but skilled, flow-dependent uncertainty estimates have been so far rarely provided. Not

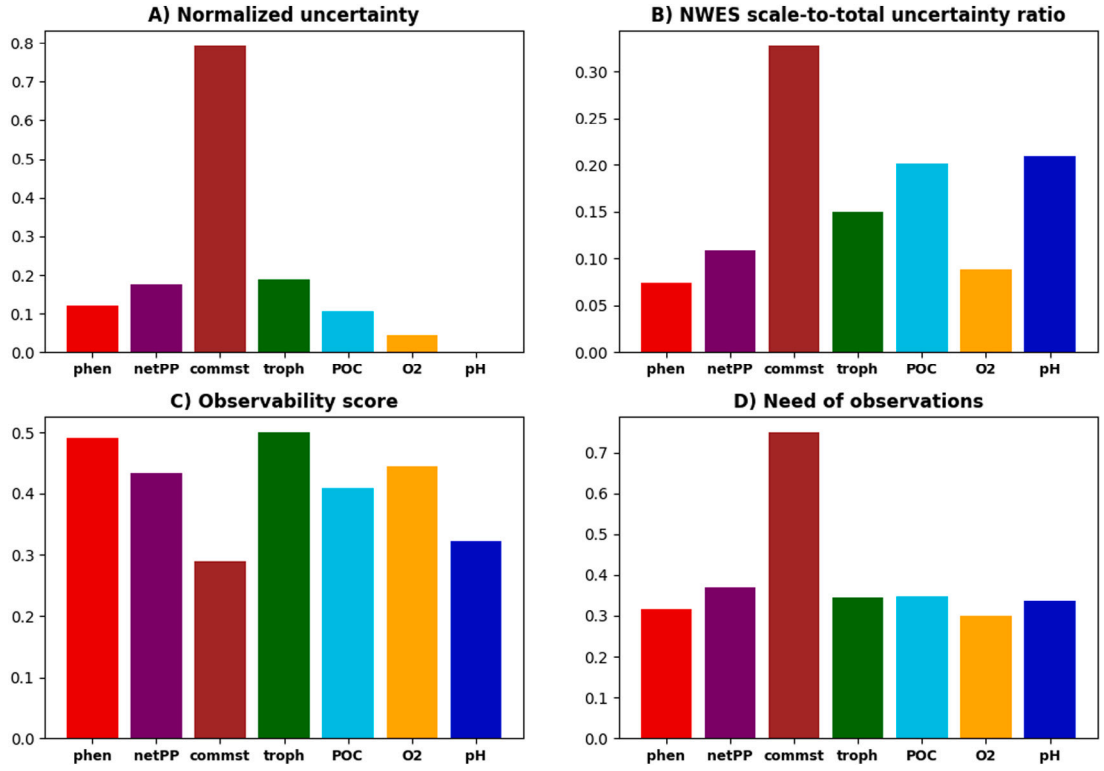

Fig. 8. The panel A shows the uncertainty measure $\mathrm{Un}(v,x,t)$ averaged across the NWES and the March-September 2018 period (metric from Eq. (5)). The panel C shows the same averages for the observability $\mathrm{Obs}(v,x,t)$ (metric from Eq. (3)) and the panel D shows the same for the need for observations score $\mathrm{NO}(v,x,t)$ (metric from Eq. (10)). The panel B shows the March-September 2018 average for the $\mathrm{Sc}(v,t)$ metric (Eq. (6)) demonstrating the proportion of uncertainty retained on the NWES scale. The abbreviations on the x-axis are: "phen": phenology, "netPP": net primary production, 'commst": phytoplankton community structure, "troph": trophic efficiency, "POC": POC at depth, "O2": bottom oxygen.

every source of MEHI uncertainty has been included in this work, but many major sources have been represented, like the biogeochemical model parameters, physical mixing scheme and the atmospheric forcing uncertainty. The advantage of these uncertainty sources is that they can be represented on the NWES by runs with relatively short (O(1) months) ensemble spin-up. We have assessed the uncertainty estimates through standard metrics and have shown the biogeochemistry uncertainties assessed through chlorophyll are reliable. In addition, we have shown that the estimated uncertainties of chlorophyll are consistent with an independent study using diagnostic methods (Fowler et al., 2023).

We have demonstrated that for all the MEHIs except for phytoplankton community structure, the MEHI uncertainty is always small compared to the typical range of variability in the daily maps produced by the NWES operational system. This indicates small noise-to-signal ratio, so apart from phytoplankton community structure, the MEHI daily products can be trusted. We have also shown that if one is interested only in the MEHI NWES mean values, the MEHI uncertainty reduces by 80%–95%. This is again true about all the MEHIs with the exception of phytoplankton community structure, where it reduces by 70%. If also time-averaging was performed on MEHI (across the simulation period), uncertainty would be further reduced by 50%–60%. These investigations might be relevant to global scale/climate applications, as well as to users interested in regional rather than site-specific assessments on shorter time scales.

Besides uncertainties, we have assessed the observability of the selected MEHIs through the total chlorophyll observations (biogeochemical variable that is most typically estimated from the numerous ocean colour and fluorescence observations). We pointed out which MEHIs and what areas of the NWES should be prioritized in future monitoring efforts. The most uncertain and least observable (from total chlorophyll) MEHI turned out to be the phytopankton community structure, highlighting the need to observe it especially in the waters near the UK and Ireland. Fortunately major progress has been achieved in the last decade in developing remotely sensed phytoplankton community structure products (e.g. Brewin et al., 2010; Brotas et al., 2013; Brewin et al., 2017), which have been assimilated into models (Ciavatta et al., 2018; Skákala et al., 2018; Pradhan et al., 2020). This work advocates for sustaining and expanding those activities in the future.

This work leads to the following recommendations to estimate MEHIs operationally: (i) beyond the usual developments of operational models (e.g. improving the model, observation and data assimilation components), we need to focus also on representing the model uncertainty. This includes developing priors for new sources of uncertainties, not represented in this work, like riverine discharge data for biogeochemistry, or spatio-temporal variations in biogeochemical model parameters. (ii) The uncertainties developed in this work were implemented in an ensemble-variational (3DEnVar) system, but longer simulations and larger amounts of independent observations are needed to better compare its performance with the existing variational (3DVar) system. (iii) The SST rank histogram analysis revealed that the physical model perturbations adopted here might need further development to better account for spatial and temporal variability on the NWES. (iv) A truly substantial source of uncertainty (and also bias) is the model structural deficiencies, e.g. lack of processes included in the model (such as mixotrophy, or predation of zooplankton by higher-trophic level species), limited capability to resolve biological species (e.g. phytoplankton types), or limited spatial resolution. It is challenging to assess those uncertainties (e.g. multi-model ensembles are computationally expensive), but perhaps steps can be taken either through stochastic spatio-temporal parameter variability, or through other, innovative techniques like multi-resolution ensembles (e.g. Jones and

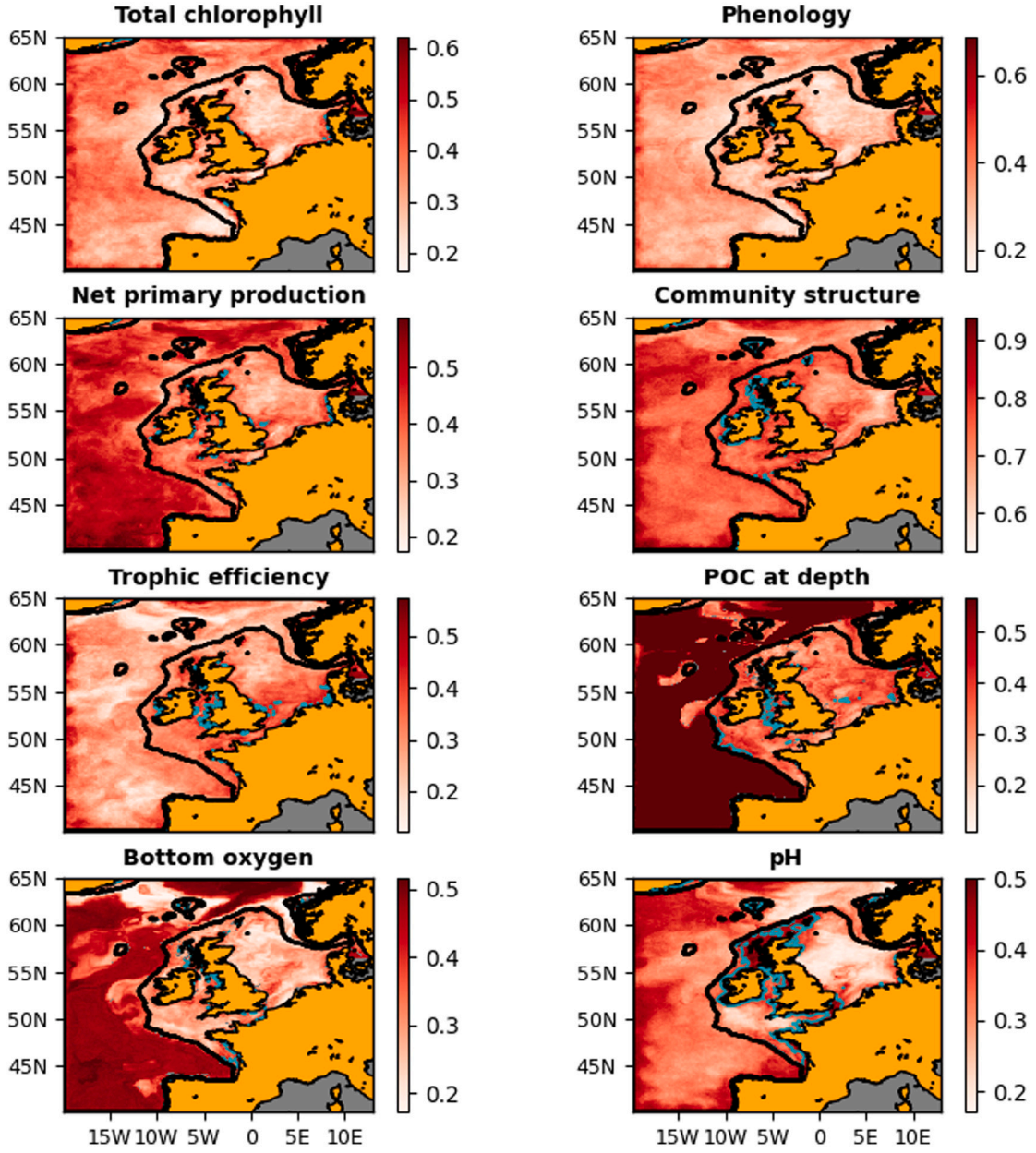

Fig. 9. The need for observations score (Eq. (10)) for the different MEHI's, averaged for the March–September 2018 period. The purpose of the Figure is to highlight the locations on the NWES where observations might be most needed. These are highlighted in cyan. The locations were determined as follows: if we take the full interval of MEHI (v) $\langle \mathrm{NO}(v,x,t)\rangle_t$ values on the NWES to be I(v) = $[\min_x(\langle \mathrm{NO}(v,x,t)\rangle_t), \max_x(\langle \mathrm{NO}(v,x,t)\rangle_t)]$, ($\min_x$, $\max_x$ are minima and maxima across NWES and $\langle\rangle_t$ is time-averaging) then the locations marked in cyan are where $\langle \mathrm{NO}(v,x,t)\rangle_t$ value is in the upper 25% of the interval I(v).

Wang, 2023) deterministic/stochastic emulators (Leeds et al., 2013; Schartau et al., 2017; Skakala et al., 2023), stochastic (e.g. multifractal) scaling models (e.g. Skákala et al., 2019), or structurally perturbed ensembles (Anugerahanti et al., 2018, 2020).

CRediT authorship contribution statement

Jozef Skákala: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **David Ford:** Methodology, Software, Writing – review & editing. **Alison Fowler:** Validation, Writing – review & editing, Visualization. **Dan Lea:** Software, Methodology. **Matthew J. Martin:** Software, Writing – review & editing, Methodology. **Stefano Ciavatta:** Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation programme, project Services based on Ecosystem data AssiMiLation: Essential Science and Solutions (SEAMLESS), under agreement number 101004032. We also acknowledge support

from the UK Natural Environment Research Council (NERC) single centre national capability programme – Climate Linked Atlantic Sector Science (NE/R015953/1) and NERC supported National Centre for Earth Observation (NCEO), contract number PR140015.

The ocean color data were provided by the European Space Agency Climate Initiative "Ocean Color" (https://esa-oceancolour-cci.org/) and the glider data assimilated here were produced as part of UK NERC funded project Alternative Framework to Assess Marine Ecosystem Functioning in Shelf Seas (AlterECO, http://projects.noc.ac.uk/altereco/), grant no. NE/P013899/1. We also used L4 observations provided by the Western Channel Observatory (https://www.westernchannelobservatory.org.uk/) and climatology data from North Sea Biogeochemical Climatology (NSBC) project (https://www.cen.uni-hamburg.de/en/icdc/data/ocean/nsbc.html). The model was forced by the atmospheric ERA5 ensemble product of The European Centre for Medium-Range Weather Forecasts (ECMWF, https://www.ecmwf.int/). The river forcing data used by the model were prepared by Sonja van Leeuwen and Helen Powley as part of UK Shelf Seas Biogeochemistry programme (contract no. NE/K001876/1) of the NERC and the Department for Environment Food and Rural Affairs (DEFRA).

We acknowledge use of the Monsoon2 system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, a strategic partnership between the Met Office and the NERC. The different outputs of simulations are stored on the Monsoon2 storage facility MASS and can be obtained upon request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.pocean.2024.103249.

References

Anugerahanti, P., Roy, S., Haines, K., 2018. A perturbed biogeochemistry model ensemble evaluated against in situ and satellite observations. Biogeosciences 15 (21), 6685–6711.

Anugerahanti, P., Roy, S., Haines, K., 2020. Perturbed biology and physics signatures in a 1-D ocean biogeochemical model ensemble. Front. Mar. Sci. 7, 549.

Artioli, Y., Blackford, J.C., Butenschön, M., Holt, J.T., Wakelin, S.L., Thomas, H., Borges, A.V., Allen, J.I., 2012. The carbonate system in the North Sea: Sensitivity and model validation. J. Mar. Syst. 102, 1–13.

Bannister, R., 2017. A review of operational methods of variational and ensemble-variational data assimilation. Q. J. R. Meteorol. Soc. 143 (703), 607–633.

Baretta, J., Ebenhöh, W., Ruardij, P., 1995. The European regional seas ecosystem model, a complex marine ecosystem model. Netherlands J. Sea Res. 33 (3–4), 233–246.

Baretta-Bekker, J., Baretta, J., Ebenhöh, W., 1997. Microbial dynamics in the marine ecosystem model ERSEM II with decoupled carbon assimilation and nutrient uptake. J. Sea Res. 38 (3–4), 195–211.

Blackford, J., 1997. An analysis of benthic biological dynamics in a North Sea ecosystem model. J. Sea Res. 38 (3–4), 213–230.

Bonan, G.B., Doney, S.C., 2018. Climate, ecosystems, and planetary futures: The challenge to predict life in earth system models. Science 359 (6375), eaam8328.

Borges, A., Schiettecatte, L.-S., Abril, G., Delille, B., Gazeau, F., 2006. Carbon dioxide in European coastal waters. Estuar. Coast. Shelf Sci. 70 (3), 375–387.

Bowler, N.E., 2008. Accounting for the effect of observation errors on verification of MOGREPS. Meteorol. Appl.: J. Forecast., Pract. Appl., Train. Tech. Model. 15 (1), 199–205.

Bowler, N., Clayton, A., Jardak, M., Lee, E., Lorenc, A., Piccolo, C., Pring, S., Wlasak, M., Barker, D., Inverarity, G., et al., 2017. Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. Q. J. R. Meteorol. Soc. 143 (704), 1280–1302.

Brewin, R.J., Ciavatta, S., Sathyendranath, S., Jackson, T., Tilstone, G., Curran, K., Airs, R.L., Cummings, D., Brotas, V., Organelli, E., et al., 2017. Uncertainty in ocean-color estimates of chlorophyll for phytoplankton groups. Front. Mar. Sci. 4, 104.

Brewin, R.J., Sathyendranath, S., Hirata, T., Lavender, S.J., Barciela, R.M., Hardman-Mountford, N.J., 2010. A three-component model of phytoplankton size class for the atlantic ocean. Ecol. Model. 221 (11), 1472–1483.

Brewin, R.J., Sathyendranath, S., Platt, T., Bouman, H., Ciavatta, S., Dall'Olmo, G., Dingle, J., Groom, S., Jönsson, B., Kostadinov, T.S., et al., 2021. Sensing the ocean biological carbon pump from space: A review of capabilities, concepts, research gaps and future developments. Earth-Sci. Rev. 217, 103604.

Brotas, V., Brewin, R.J., Sá, C., Brito, A.C., Silva, A., Mendes, C.R., Diniz, T., Kaufmann, M., Tarran, G., Groom, S.B., et al., 2013. Deriving phytoplankton size classes from satellite data: Validation along a trophic gradient in the eastern Atlantic Ocean. Remote Sens. Environ. 134, 66–77.

Bruggeman, J., Bolding, K., 2014. A general framework for aquatic biogeochemical models. Environ. Model. Softw. 61, 249–265.

Bruggeman, J., Bolding, K., 2020. Framework for aquatic biogeochemical models. http://dx.doi.org/10.5281/zenodo.3817997.

Burgers, G., Van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126 (6), 1719–1724.

Butenschön, M., Clark, J., Aldridge, J.N., Allen, J.I., Artioli, Y., Blackford, J., Bruggeman, J., Cazenave, P., Ciavatta, S., Kay, S., et al., 2016. ERSEM 15.06: A generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels. Geosci. Model Dev. 9 (4), 1293–1339.

Campbell, J.W., 1995. The lognormal distribution as a model for bio-optical variability in the sea. J. Geophys. Res.: Oceans 100 (C7), 13237–13254.

Casati, B., Wilson, L., Stephenson, D., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E., Brown, B., Mason, S., 2008. Forecast verification: current status and future directions. Meteorol. Appl.: J. Forecast., Pract. Appl., Train. Tech. Model. 15 (1), 3–18.

Ciavatta, S., Brewin, R., Skakala, J., Polimene, L., de Mora, L., Artioli, Y., Allen, J.I., 2018. Assimilation of ocean-color plankton functional types to improve marine ecosystem simulations. J. Geophys. Res.: Oceans 123 (2), 834–854.

Ciavatta, S., Kay, S., Saux-Picart, S., Butenschön, M., Allen, J., 2016. Decadal reanalysis of biogeochemical indicators and fluxes in the north west European shelf-sea ecosystem. J. Geophys. Res.: Oceans 121 (3), 1824–1845.

Ciavatta, S., Torres, R., Saux-Picart, S., Allen, J.I., 2011. Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? J. Geophys. Res.: Oceans 116 (C12).

Claustre, H., Johnson, K.S., Takeshita, Y., 2020. Observing the global ocean with biogeochemical-argo. Annu. Rev. Mar. Sci. 12, 23–48.

Der Kiureghian, A., Ditlevsen, O., 2009. Aleatory or epistemic? Does it matter? Struct. Saf. 31 (2), 105–112.

Desroziers, G., Berre, L., Chapnik, B., Poli, P., 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R. Meteorol. Soc.: J. Atmos. Sci., Appl. Meteorol. Phys. Oceanogr. 131 (613), 3385–3396.

Doron, M., Brasseur, P., Brankart, J.-M., 2011. Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical–biogeochemical model: Twin experiments. J. Mar. Syst. 87 (3–4), 194–207.

Doron, M., Brasseur, P., Brankart, J.-M., Losa, S.N., Melet, A., 2013. Stochastic estimation of biogeochemical parameters from globcolour ocean colour satellite data in a north atlantic 3D ocean coupled physical–biogeochemical model. J. Mar. Syst. 117, 81–95.

Dowd, M., 2007. Bayesian statistical data assimilation for ecosystem models using Markov chain Monte Carlo. J. Mar. Syst. 68 (3–4), 439–456.

Evensen, G., 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dyn. 53, 343–367.

Fennel, K., Gehlen, M., Brasseur, P., Brown, C.W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C.A., Ford, D., Friedrichs, M.A., et al., 2019. Advancing marine biogeochemical and ecosystem reanalyses and forecasts as tools for monitoring and managing ecosystem health. Front. Mar. Sci. 6, 89.

Fennel, K., Mattern, J.P., Doney, S.C., Bopp, L., Moore, A.M., Wang, B., Yu, L., 2022. Ocean biogeochemical modelling. Nature Rev. Methods Primers 2 (1), 76.

Flynn, K.J., Torres, R., Irigoien, X., Blackford, J.C., 2022. Plankton digital twins—A new research tool. J. Plankton Res..

Fontana, C., Grenz, C., Pinazo, C., Marsaleix, P., Diaz, F., 2009. Assimilation of SeaWiFS chlorophyll data into a 3D-coupled physical–biogeochemical model applied to a freshwater-influenced coastal zone. Cont. Shelf Res. 29 (11–12), 1397–1409.

Ford, D., 2021. Assimilating synthetic biogeochemical-argo and ocean colour observations into a global ocean model to inform observing system design. Biogeosciences 18 (2), 509–534.

Ford, D.A., Grossberg, S., Rinaldi, G., Menon, P.P., Palmer, M.R., Skákala, J., Smyth, T., Williams, C.A., Lorenzo Lopez, A., Ciavatta, S., 2022. A solution for autonomous, adaptive monitoring of coastal ocean ecosystems: Integrating ocean robots and operational forecasts. Front. Mar. Sci. 9, 1067174.

Ford, D.A., van der Molen, J., Hyder, K., Bacon, J., Barciela, R., Creach, V., McEwan, R., Ruardij, P., Forster, R., 2017. Observing and modelling phytoplankton community structure in the North Sea. Biogeosciences 14 (6), 1419–1444.

Fowler, A.M., Skákala, J., Ford, D., 2023. Validating and improving the uncertainty assumptions for the assimilation of ocean-colour-derived chlorophyll a into a marine biogeochemistry model of the Northwest European shelf seas. Q. J. R. Meteorol. Soc. 149 (750), 300–324.

Friedrich, T., Timmermann, A., Abe-Ouchi, A., Bates, N., Chikamoto, M., Church, M., Dore, J., Gledhill, D., Gonzalez-Davila, M., Heinemann, M., et al., 2012. Detecting regional anthropogenic trends in ocean acidification against natural variability. Nature Clim. Change 2 (3), 167–171.

Garnier, F., Brankart, J.-M., Brasseur, P., Cosme, E., 2016. Stochastic parameterizations of biogeochemical uncertainties in a 1/4° NEMO/PISCES model for probabilistic comparisons with ocean color data. J. Mar. Syst. 155, 59–72.

Garrett, C., 1992. Oceanographic and modelling considerations in marine environmental protection. Mar. Pollut. Bull. 25 (1–4), 41–44.
Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D., et al., 2015. Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments. J. Oper. Oceanogr. 8 (sup1), s168–s187.
Geider, R., MacIntyre, H., Kana, T., 1997. Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature. Mar. Ecol. Prog. Ser. 148, 187–200.
Germineaud, C., Brankart, J.-M., Brasseur, P., 2019. An ensemble-based probabilistic score approach to compare observation scenarios: an application to biogeochemical-argo deployments. J. Atmos. Ocean. Technol. (2019).
Gharamti, M., Samuelsen, A., Bertino, L., Simon, E., Korosov, A., Daewel, U., 2017a. Online tuning of ocean biogeochemical model parameters using ensemble estimation techniques: Application to a one-dimensional model in the north atlantic. J. Mar. Syst. 168, 1–16.
Gharamti, M., Tjiputra, J., Bethke, I., Samuelsen, A., Skjelvan, I., Bentsen, M., Bertino, L., 2017b. Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. Ocean Model. 112, 65–89.
Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. Mon. Weather Rev. 129 (3), 550–560.
Harris, R., 2010. The L4 time-series: the first 20 years. J. Plankton Res. 32 (5), 577–583.
Hinrichs, I., Gouretski, V., Pätch, J., Emeis, K., Stammer, D., 2017. North Sea Biogeochemical Climatology. Universität Hamburg.
Hollingsworth, A., Lönnberg, P., 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. Tellus A 38 (2), 111–136.
Hüllermeier, E., Waegeman, W., 2021. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Mach. Learn. 110, 457–506.
Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., Raynaud, L., 2010. Ensemble of Data Assimilations at ECMWF. ECMWF, Reading, UK.
Jahnke, R.A., 2010. Global synthesis. In: Carbon and Nutrient Fluxes in Continental Margins. Springer, pp. 597–615.
Jones, E.A., Wang, X., 2023. A multiresolution ensemble hybrid 4DEnVar with variable ensemble sizes to improve global and tropical cyclone track numerical prediction. Mon. Weather Rev. 151 (5), 1145–1166.
Kalman, R., 1959. On the general theory of control systems. IRE Trans. Autom. Control 4 (3), 110.
Kalman, R.E., 1963. Mathematical description of linear dynamical systems. J. Soc. Ind. Appl. Math., Ser. A: Control 1 (2), 152–192.
Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press.
Kang, W., Xu, L., 2009. Computational analysis of control systems using dynamic optimization. arXiv preprint arXiv:0906.0215.
Kay, S., McEwan, R., Ford, D., 2019. North west European shelf production centre Northwestshelf_analysis_forecast_bio_004_011, quality information document.
King, S., Kang, W., Xu, L., 2015. Observability for optimal sensor locations in data assimilation. Int. J. Dyn. Control 3, 416–424.
King, S., Kang, W., Xu, L., Baker, N.L., 2017. Information quantification for data assimilation. In: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, vol. III, Springer, pp. 121–139.
King, R.R., While, J., Martin, M.J., Lea, D.J., Lemieux-Dudon, B., Waters, J., O'Dea, E., 2018. Improving the initialisation of the met office operational shelf-seas model. Ocean Model. 130, 1–14.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.
Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J.R., Dunne, J.P., Gehlen, M., Ilyina, T., John, J.G., et al., 2020. Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. Biogeosciences 17 (13), 3439–3470.
Lea, D.J., While, J., Martin, M.J., Weaver, A., Storto, A., Chrust, M., 2022. A new global ocean ensemble system at the met office: Assessing the impact of hybrid data assimilation and inflation settings. Q. J. R. Meteorol. Soc. 148 (745), 1996–2030.
Leeds, W., Wikle, C., Fiechter, J., Brown, J., Milliff, R., 2013. Modeling 3-D spatio-temporal biogeochemical processes with a forest of 1-D statistical emulators. Environmetrics 24 (1), 1–12.
Lenhart, H.-J., Mills, D.K., Baretta-Bekker, H., Van Leeuwen, S.M., van der Molen, J., Baretta, J.W., Blaas, M., Desmit, X., Kühn, W., Lacroix, G., et al., 2010. Predicting the consequences of nutrient reduction on the eutrophication status of the North Sea. J. Mar. Syst. 81 (1–2), 148–170.
Lorenc, A.C., Jardak, M., 2018. A comparison of hybrid variational data assimilation methods for global NWP. Q. J. R. Meteorol. Soc. 144 (717), 2748–2760.
Loveday, B.R., Smyth, T., Akpinar, A., Hull, T., Inall, M.E., Kaiser, J., Queste, B.Y., Tobermann, M., Williams, C.A., Palmer, M.R., 2022. Application of a new net primary production methodology: A daily to annual-scale data set for the North Sea, derived from autonomous underwater gliders and satellite earth observation. Earth Syst. Sci. Data 14 (9), 3997–4016.
Madec, G., et al., 2015. NEMO Ocean Engine. Institut Pierre-Simon Laplace.
Marine Systems Modelling Group, P.M.L., 2020. European regional seas ecosystem model. http://dx.doi.org/10.5281/zenodo.3817997.
Martinelli, A., 2022. Extension of the observability rank condition to time-varying nonlinear systems. IEEE Trans. Automat. Control 67 (9), 5002–5008.
Meier, H.M., Edman, M., Eilola, K., Placke, M., Neumann, T., Andersson, H.C., Brunnabend, S.-E., Dieterich, C., Frauen, C., Friedland, R., et al., 2019. Assessment of uncertainties in scenario simulations of biogeochemical cycles in the Baltic Sea. Front. Mar. Sci. 6, 46.
Mogensen, K., Balmaseda, M., Weaver, A., Martin, M., Vidard, A., 2009. NEMOVAR: A variational data assimilation system for the NEMO ocean model. ECMWF Newsl. 120, 17–22.
Mogensen, K., Balmaseda, M.A., Weaver, A., et al., 2012. The NEMOVAR ocean data assimilation system as implemented in the ecmwf ocean analysis for system 4.
Moore, B., 1981. Principal component analysis in linear systems: Controllability, observability, and model reduction. IEEE Trans. Autom. Control 26 (1), 17–32.
O'Dea, E., Furner, R., Wakelin, S., Siddorn, J., While, J., Sykes, P., King, R., Holt, J., Hewitt, H., 2017. The CO5 configuration of the 7 km Atlantic Margin Model: large-scale biases and sensitivity to forcing, physics options and vertical resolution. Geosci. Model Dev. 10 (8), 2947.
O'Neill, B.C., Tebaldi, C., Van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., et al., 2016. The scenario model intercomparison project (scenariomip) for CMIP6. Geosci. Model Dev. 9 (9), 3461–3482.
Palmer, T., 2019. The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. Q. J. R. Meteorol. Soc. 145, 12–24.
Pauly, D., Christensen, V., Guénette, S., Pitcher, T.J., Sumaila, U.R., Walters, C.J., Watson, R., Zeller, D., 2002. Towards sustainability in world fisheries. Nature 418 (6898), 689.
Payne, M.R., Barange, M., Cheung, W.W., MacKenzie, B.R., Batchelder, H.P., Cormon, X., Eddy, T.D., Fernandes, J.A., Hollowed, A.B., Jones, M.C., et al., 2016. Uncertainties in projecting climate-change impacts in marine ecosystems. ICES J. Mar. Sci. 73 (5), 1272–1282.
Pradhan, H.K., Völker, C., Losa, S.N., Bracher, A., Nerger, L., 2020. Global assimilation of ocean-color data of phytoplankton functional types: Impact of different data sets. J. Geophys. Res.: Oceans 125 (2), e2019JC015586.
Rodwell, M.J., Lang, S.T.K., Ingleby, N.B., Bormann, N., Hólm, E., Rabier, F., Richardson, D.S., Yamaguchi, M., 2016. Reliability in ensemble data assimilation. Q. J. R. Meteorol. Soc. 142 (694), 443–454. http://dx.doi.org/10.1002/qj.2663, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2663, arXiv:https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2663.
Roulston, M.S., Smith, L.A., 2002. Evaluating probabilistic forecasts using information theory. Mon. Weather Rev. 130 (6), 1653–1660.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: the Primer. John Wiley & Sons.
Saltelli, A., Tarantola, S., Campolongo, F., 2000. Sensitivity analysis as an ingredient of modeling. Stat. Sci. 377–395.
Sathyendranath, S., Brewin, R.J., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., Cipollini, P., Couto, A.B., Dingle, J., Doerffer, R., et al., 2019. An ocean-colour time series for use in climate studies: The experience of the ocean-colour climate change initiative (OC-CCI). Sensors 19 (19), 4285.
Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B.A., Slawig, T., Oschlies, A., 2017. Reviews and syntheses: parameter identification in marine planktonic ecosystem modelling. Biogeosciences 14 (6), 1647–1701.
Schmidtko, S., Stramma, L., Visbeck, M., 2017. Decline in global oceanic oxygen content during the past five decades. Nature 542 (7641), 335–339.
Siddorn, J., Furner, R., 2013. An analytical stretching function that combines the best attributes of geopotential and terrain-following vertical coordinates. Ocean Model. 66, 1–13.
Skakala, J., Awty-Carroll, K., Menon, P.P., Wang, K., Lessin, G., 2023. Future digital twins: emulating a highly complex marine biogeochemical model with machine learning to predict hypoxia. Front. Mar. Sci. 10, 1058837.
Skákala, J., Bruggeman, J., Brewin, R.J., Ford, D.A., Ciavatta, S., 2020. Improved representation of underwater light field and its impact on ecosystem dynamics: A study in the North Sea. J. Geophys. Res.: Oceans e2020JC016122.
Skákala, J., Bruggeman, J., Ford, D., Wakelin, S., Akpınar, A., Hull, T., Kaiser, J., Loveday, B.R., O'Dea, E., Williams, C.A., et al., 2022. The impact of ocean biogeochemistry on physics and its consequences for modelling shelf seas. Ocean Model. 172, 101976.
Skákala, J., Ford, D., Brewin, R.J., McEwan, R., Kay, S., Taylor, B., de Mora, L., Ciavatta, S., 2018. The assimilation of phytoplankton functional types for operational forecasting in the northwest European shelf. J. Geophys. Res.: Oceans 123 (8), 5230–5247.
Skákala, J., Ford, D.A., Bruggeman, J., Hull, T., Kaiser, J., King, R.R., Loveday, B.R., Palmer, M.R., Smyth, T.J., Williams, C.A.J., Ciavatta, S., 2021. Towards a multi-platform assimilative system for ocean biogeochemistry. Earth Space Sci. Open Arch. ESSOAr, JGR-Oceans submitted for publication.
Skákala, J., Smyth, T.J., Torres, R., Buckingham, C.E., Brearley, A., Hyder, P., Coward, A.C., 2019. SST dynamics at different scales: Evaluating the oceanographic model resolution skill to represent SST processes in the southern ocean. J. Geophys. Res.: Oceans 124 (4), 2546–2570.

Slingo, J., Palmer, T., 2011. Uncertainty in weather and climate prediction. Phil. Trans. R. Soc. A 369 (1956), 4751–4767.

Stefano, C., Lazzari, P., Alvarez Suarez, E., Jorn, B., Capet, A., Cossarini, G., Daryabor, F., Lars, N., Jozef, S., Teruzzi, A., et al., 2022. Observability of the target indicators and parameter sensitivity in the 1D CMEMS sites. Deliverable report (d3. 2) of project H2020 SEAMLESS. In: Observability of the Target Indicators and Parameter Sensitivity in the 1D CMEMS Sites. Deliverable Report (D3. 2) of Project H2020 SEAMLESS.

Stigter, J., Joubert, D., Molenaar, J., 2017. Observability of complex systems: Finding the gap. Sci. Rep. 7 (1), 16566.

Storkey, D., Blockley, E., Furner, R., Guiavarc'h, C., Lea, D., Martin, M., Barciela, R., Hines, A., Hyder, P., Siddorn, J., 2010. Forecasting the ocean state using NEMO: The new FOAM system. J. Oper. Oceanogr. 3 (1), 3–15.

Storto, A., Andriopoulos, P., 2021. A new stochastic ocean physics package and its application to hybrid-covariance data assimilation. Q. J. R. Meteorol. Soc. 147 (736), 1691–1725.

Telszewski, M., Palacz, A., Fischer, A., 2018. Biogeochemical in situ observations–motivation, status, and new frontiers. New Front. Oper. Oceanogr. 131–160.

Van Leeuwen, P.J., 2020. A consistent interpretation of the stochastic version of the ensemble Kalman filter. Q. J. R. Meteorol. Soc. 146 (731), 2815–2825.

Wang, Z.A., Moustahfid, H., Mueller, A.V., Michel, A.P., Mowlem, M., Glazer, B.T., Mooney, T.A., Michaels, W., McQuillan, J.S., Robidart, J.C., et al., 2019. Advancing observation of ocean biogeochemistry, biology, and ecosystems with cost-effective in situ sensing technologies. Front. Mar. Sci. 519.

Waters, J., Lea, D.J., Martin, M.J., Mirouze, I., Weaver, A., While, J., 2015. Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. Q. J. R. Meteorol. Soc. 141 (687), 333–349.

Weaver, A., Chrust, M., Ménétrier, B., Piacentini, A., Tshimanga, J., Yang, Y., Gürol, S., Zuo, H., 2018. Using ensemble-estimated background error variances and correlation scales in the NEMOVAR system. In: Report TR/PA/18/15. CERFACS, Toulouse, France, p. 39.

Wilson, J.D., Andrews, O., Katavouta, A., de Melo Viríssimo, F., Death, R.M., Adloff, M., Baker, C.A., Blackledge, B., Goldsworth, F.W., Kennedy-Asser, A.T., et al., 2022. The biological carbon pump in CMIP6 models: 21st century trends and uncertainties. Proc. Natl. Acad. Sci. 119 (29), e2204369119.

Zedler, S.E., Polton, J.A., King, R.R., Wakelin, S.L., 2023. The effect of uncertain river forcing on the thermohaline properties of the north west European shelf seas. Ocean Model. 183, 102196.

Zuo, H., Balmaseda, M., De Boisseson, E., Hirahara, S., Chrust, M., De Rosnay, P., 2017. A Generic Ensemble Generation Scheme for Data Assimilation and Ocean Analysis. European Centre for Medium Range Weather Forecasts.