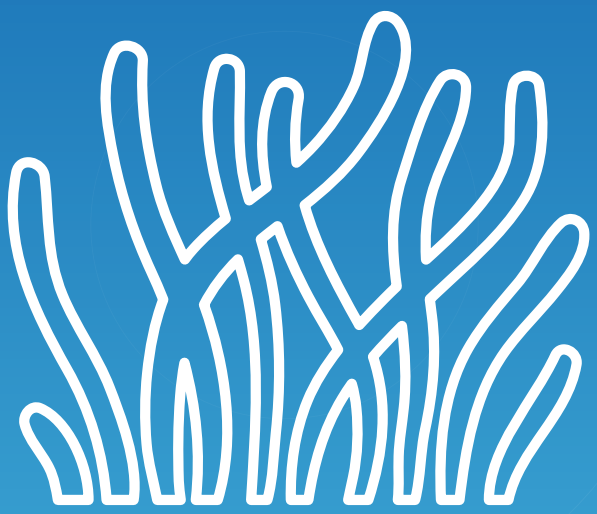
9.
Biogeochemical modelling

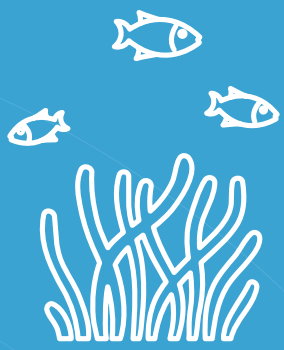


CHAPTER COORDINATOR

Elodie Gutknecht

CHAPTER AUTHORS *(in alphabetical order)*

Laurent Bertino, Pierre Brasseur, Stefano Ciavatta, Gianpiero Cossarini, Katja Fennel, David Ford, Marilaure Grégoire, Diane Lavoie, and Patrick Lehodey



9. Biogeochemical modelling

9.1. General introduction to Biogeochemical models

9.1.1. Objective, applications and beneficiaries

9.1.2. Fundamental theoretical background

9.1.2.1. Biogeochemical modelling

9.1.2.2. Model calibration

9.1.2.3. Physical-Biogeochemical coupling

9.1.2.4. From open ocean to coastal ecosystems

9.1.2.5. Potential predictability of ocean biogeochemistry

9.2. Biogeochemical forecast and multi-year systems

9.2.1. Architecture singularities

9.2.1.1. Physical, optical, and biogeochemical components

9.2.1.2. Propagation of uncertainties

9.2.1.3. BGC Data singularities

9.2.2. Input data: available sources and data handling

9.2.2.1. Physical conditions

9.2.2.2. Observational data

9.2.2.3. Climatologies, databases, and atlases

9.2.2.4. Atmospheric surface forcing

9.2.2.5. External inputs

9.2.2.6. Units

9.2.3. Modelling component

9.2.3.1. Numerical and discretisation choices

9.2.3.2. The different biogeochemical models

9.2.3.3. Connections Ocean-Earth systems

9.2.4. Ensemble modelling

9.2.5. Data assimilation systems

9.2.5.1. Biogeochemical state and parameter estimation

9.2.5.2. Assimilated observational products

9.2.5.3. Biogeochemical data assimilation methods

9.2.5.4. Current challenges and opportunities

Summary

Marine biogeochemistry is the study of essential chemical elements in the ocean (such as carbon, nitrogen, oxygen, and phosphorus), and of their interactions with marine organisms. Biogeochemical cycles are driven by physical transport, chemical reactions, absorption, and transformation by plankton and other organisms, which form the basis of the oceanic food web.

In the last decades, the interest for this cross-disciplinary science has greatly increased due to the occurrence of significant changes in the marine environment closely linked to the alteration of the biogeochemical cycles in the ocean. These alterations include phenomena such as acidification, coral bleaching, eutrophication, deoxygenation, harmful algal blooms, regime shifts in plankton, invasive species, and other processes deteriorating water quality and impacting the whole marine ecosystem.

Monitoring and forecasting the biogeochemical and ecosystem components of the ocean, also referred to as "Green Ocean", are essential for a better understanding of the current status and changes in ocean health and ecosystem functioning. Such operational systems provide indicators useful to scientists, industry (e.g. fisheries and aquaculture), policy makers and environmental agencies for the prediction of events, the management of living marine resources, and can support the decision-making process to respond to environmental changes.

This chapter gives an overview of the Green Ocean component of OOFS. The first section addresses the objectives, applications and beneficiaries of the Green Ocean and introduces the fundamental theoretical knowledge of marine biogeochemical modelling. The second section details and discusses each component of a biogeochemical OOFS to guide new forecasters in biogeochemistry. Modelling of higher trophic levels is introduced. Finally, several operational systems are mentioned as examples.

9.1. General introduction to Biogeochemical models

9.1.1. Objective, applications and beneficiaries

Human activities, primarily the combustion of fossil fuels, cement production, and the industrial production of nitrogen-based fertilisers, are leading to ocean warming, acidification, deoxygenation, and coastal eutrophication, thus putting ever-increasing and compounding pressures on marine ecosystems (Figure 9.1).

At the same time, the ocean is serving as a major sink of carbon dioxide (CO2), the most important anthropogenic greenhouse gas. This contributes to mitigating global warming, but the magnitude of this sink is likely to diminish. Our ability to quantify these phenomena and project their future course hinges on a mechanistic understanding of the biogeochemical cycles of carbon, oxygen, and nutrients in the ocean and how they are changing.

The Marine BGC, the study of elemental cycles and their interactions with the environment and living organisms, is a multidisciplinary science at the crossroads between ocean physics, chemistry, and biology, and intersects with atmospheric and terrestrial sciences as well as social science and environmental policy. As an example, Figure 9.2 illustrates the complex carbon cycle in the ocean and the interactions between biological, chemical, and physical processes.

Figure 9.1. Threats on marine ecosystems. Changes and alterations in the marine environment observed in recent decades include acidification, coral bleaching, eutrophication, deoxygenation, harmful algal blooms, changes in planktonic regimes, invasive species, etc.

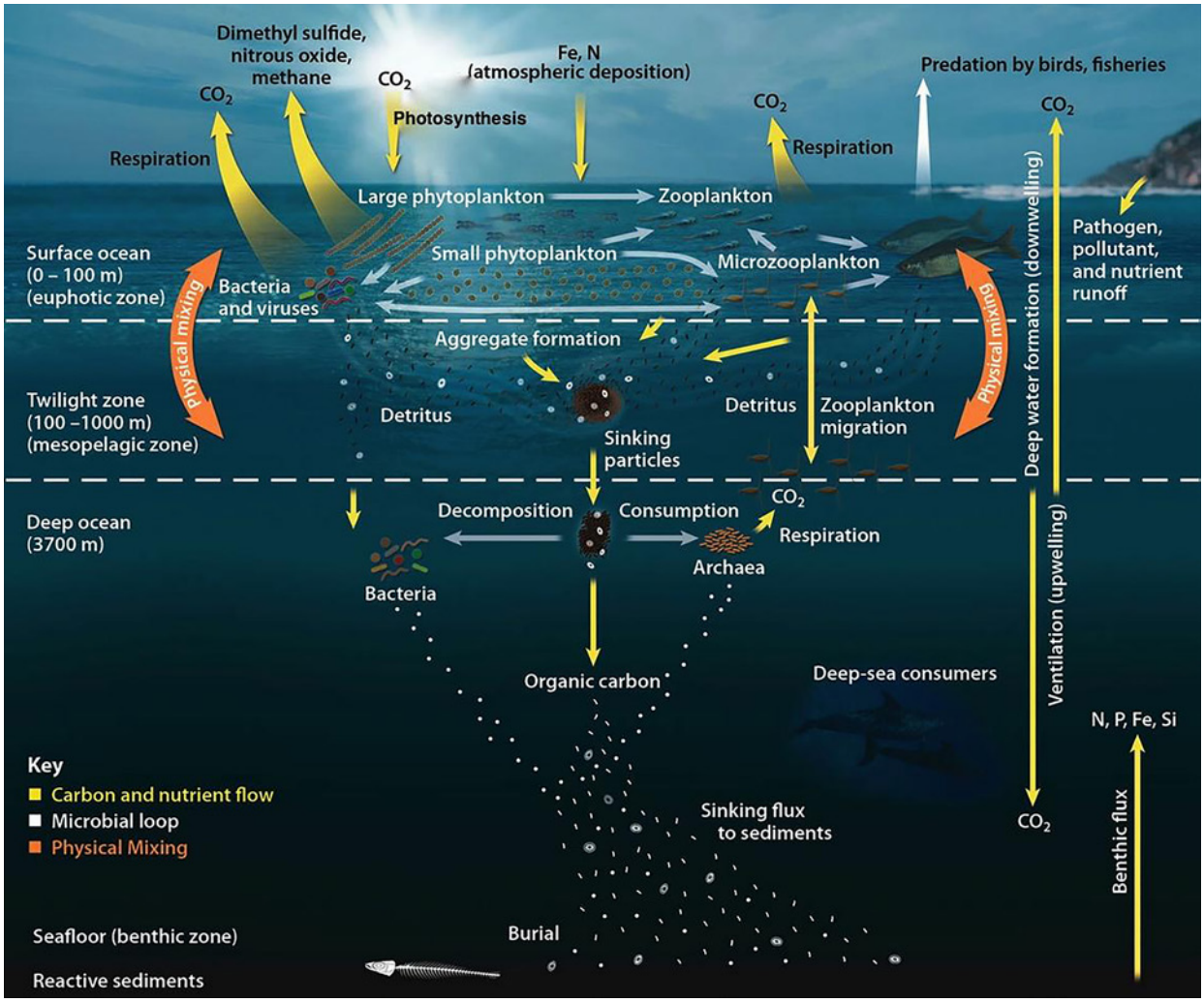


Figure 9.2. Cycling of carbon in the marine food chain. Phytoplankton assimilate CO2 via photosynthesis in the euphotic zone and are consumed by zooplankton. Zooplankton are the initial prey for many small and large aquatic organisms. Carbon is thus transferred further up the food web to higher-level predators. Different mechanisms contribute to the export and storage of carbon into the deep ocean. The carbon cycle in the ocean is complex and influenced by biological, chemical, and physical processes (credit: Oak Ridge National Laboratory at [1]).

Ocean BGC models describe the base of the marine food chain from bacteria to mesozooplankton and couple the cycles of carbon (C), nitrogen (N), oxygen (O2), phosphorus (P) and silicon (Si). They mostly focus on plankton, classifying the plankton diversity in accordance with their functional characteristics, the so-called Plankton Functional Types (PFTs). Species at higher trophic levels, such as fish and marine mammals play a lesser role in elemental cycling, they are thus generally not explicitly represented in BGC models, but they are very important for ecosystem models that focus on the ecology/biology of marine organisms. BGC and ecosystem models are sometimes referred to indistinctly because they can overlap in their representation of the lower trophic levels. Specific modelling approaches, like Lagrangian modelling, habitat modelling, or food web models, are used to connect BGC with the high trophic levels (e.g. fish).

The implementation of accurate OOFSs requires sustained, systematic, and NRT observation from (sub)mesoscale to large scale to initialise, parameterize, and validate ocean models. NRT information in operational oceanography means a description of the present situation with a delay of a few minutes to a few days.

1. https://www.ornl.gov/

The forecast of ocean physics has considerably improved in the last decades, reaching a high level of predictability (Chapter 5). The evolving equations governing the physical dynamics are based on physical laws, the model parameterizations are quite well-established, and the abundance of observations for temperature, salinity, and sea level height offers a way to improve model predictions through data assimilation. Forecasting of the Green Ocean has been developed more recently and it has not yet reached the same level of maturity, in most cases being incorporated into already existing physical OOFS. The formulation of ecosystem models is still empirical and the scarcity of in-situ biological and BGC data critically limits the capabilities to constrain their parameterization and to improve their performances through a robust data-model comparison exercise and data assimilation. The scarcity of data is even more critical in NRT, limiting data assimilation to surface chlorophyll-a (Chla) derived from satellite reflectance (Fennel et al., 2019).

The advent of in-situ robotic platforms combined with high resolution satellite products for the Green Ocean have the potential to palliate this deficiency. For instance, the advent of hyperspectral satellites is promising in terms of delivery of surface information on PFTs, detection of harmful algal blooms, and benthic habitat mapping, while the boost in robotic platforms will offer huge opportunities to map the (deep) seafloor with an unprecedented level of details. The combination of marine robotics, image analysis, machine learning, new sensor development, and the coordination of robotic platforms and satellite sensors will constitute a significant breakthrough in our knowledge of marine ecosystems. All this information would need to be integrated in models for forecasting and producing high quality reanalysis of the Green Ocean to support the production of added value products and innovative services. Coordination of Ocean OSSEs can help to design the new observing biological and biogeochemical systems with maximal impact to users, yet their development is still insufficient and should be encouraged (Le Traon et al., 2019).

Ultimately, BGC OOFS systems serve major environmental and societal issues, including the Ocean's role in the global carbon cycle and the impacts of natural changes and anthropogenic stressors in the physical-chemical marine environment on ecosystems and human activities. Applications range from multi-decadal retrospective simulations (namely, "reanalyses"), operational analysis of the current conditions ("nowcasts"), short-term and seasonal predictions ("forecasts"), scenario simulations, and climate change projections. These integrated systems are essential not only for a better understanding of the current status of key biogeochemical and ecosystem processes in the ocean and how they are changing, but also to provide stakeholders, policy makers and environmental agencies with indicators of ocean health in order to take appropriate mitigation, adaptation, conservation, and protection measures for living marine organisms and their habitats but also for human health.

"A predicted ocean whereby society has the capacity to understand current and future ocean conditions, forecast change and impact on human wellbeing and livelihoods" is an expected outcome of the United Nations Decade of Ocean Science for Sustainable Development, 2021-2030 (Ryabinin et al., 2019), supported also by the Sustainable Development Goals 14 (Life below water), 8 (Decent work and economic growth), and 9 (Industry, innovation and infrastructure).

9.1.2. Fundamental theoretical background

9.1.2.1. Biogeochemical modelling

Plankton (including phytoplankton and zooplankton) are organisms which are carried by tides and currents, or do not swim well enough to move against them. They form the base of the marine ecosystem and are a central component of the BGC models that simulate the cycling of elements through seawater and plankton.

Most models take an "NPZD" approach, simulating:

- Nutrients: substances which organisms require for growth.
- Phytoplankton: microscopic algae which obtain energy from sunlight through photosynthesis.
- Zooplankton: planktonic animals which obtain energy by eating other organisms.
- Detritus: dead and excreted organic matter.

Each of these is represented by one or more state variables, depending on the complexity of the model. Rather than considering individual organisms, state variables represent concentrations of elements such as nitrogen or carbon. They are often called tracers because they are transported and diffused by ocean dynamics.

As with physical models, BGC models are discretized on a grid covering the region of interest and require suitable initial and boundary conditions for each state variable. At each grid point, the evolution of a state variable C is given by the equation:

$$\frac{\partial C}{\partial t} = -\nabla \cdot (CU) + D^C + SMS(C) \qquad (9.1)$$

where $\nabla \cdot (CU)$ and D^C are the advection and diffusion terms equivalent to those used for temperature and salinity in physical models (please refer to Chapter 5). ∇ is the generalised derivative vector operator, t is the time, U the vector velocity, and D^C is the parameterization of small-scale physics for the tracer. The $SMS(C)$ stands for source-minus-sink terms for the tracer C and represents the BGC processes simulated by the model. Each 1D water column is normally treated independently, with lateral interactions limited to advection and diffusion. Most BGC models are formulated to conserve mass.

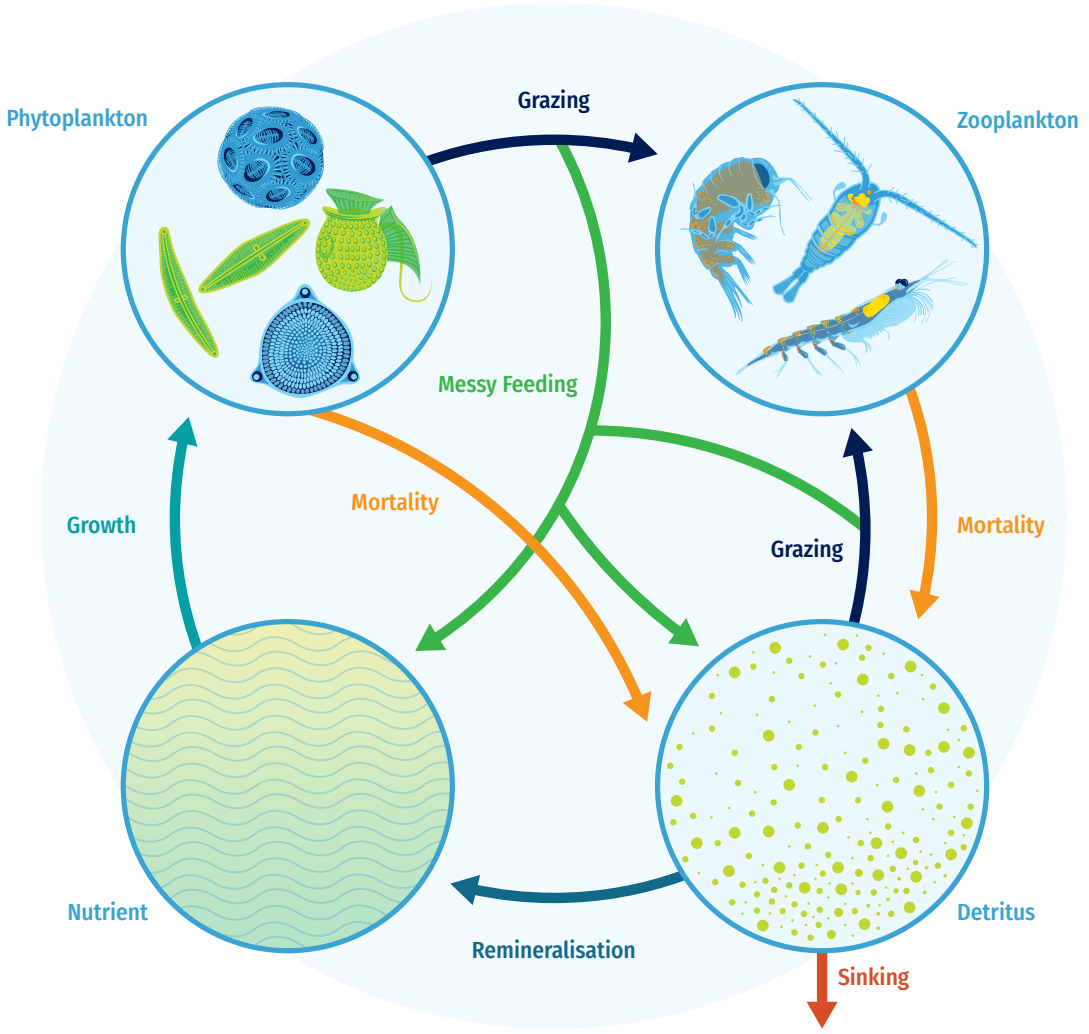


Figure 9.3. Schematic of a basic NPZD model considering four state variables, one for each compartment.

Unlike the Navier-Stokes equations for physical models (Chapter 5), there is no known set of laws defining biological behaviour. Instead, empirical relationships are used to describe observed processes such as growth and mortality.

The basic source-minus-sink terms usually modelled in a NPZD model (Figure 9.3) are:

- **Phytoplankton growth or Primary production**: the creation of organic matter through photosynthesis. It is a function of phytoplankton concentration, nutrient availability, and light availability. It can also be regulated by temperature.
- **Grazing**: zooplankton eating phytoplankton and detritus.
- **Mortality**: death through natural causes, e.g. viruses, predation by higher trophic levels (fish and marine mammals), etc.
- **Messy feeding**: zooplankton graze inefficiently, and a proportion of organic matter enters the nutrient or detritus pool rather than being ingested by zooplankton.
- **Remineralisation**: bacteria break down the organic matter in detritus, which is converted back to nutrients.
- **Sinking**: detritus sinks through the water column due to gravity.

In this case, the differential equations for phytoplankton (P), zooplankton (Z), detritus (D), and nutrients (N) are as follows:

$$\frac{\partial P}{\partial t} = \mu_P P - G_P Z - m_P P \tag{9.2}$$

where phytoplankton evolution depends on primary production, grazing and mortality;

$$\frac{\partial Z}{\partial t} = (\alpha_D + \alpha_N)(G_P + G_D)Z - m_Z Z \tag{9.3}$$

where zooplankton evolution depends on grazing and mortality;

$$\frac{\partial D}{\partial t} = m_P P + m_Z Z - G_D Z + (1-\alpha_D)(G_P + G_D)Z - rem_D D - w_D D \quad (9.4)$$

where detritus evolution depends on mortality, grazing, messy feeding, remineralisation and sinking;

$$\frac{\partial N}{\partial t} = \mu_P P + (1-\alpha_N)(G_P + G_D)Z + rem_D D \quad (9.5)$$

where nutrients evolution depends on primary production, messy feeding, and remineralisation.

μ_P is the growth rate of phytoplankton due to photosynthesis; m_P and m_z are the mortality rates of phytoplankton and zooplankton; G_P and G_D are the grazing rates of zooplankton on phytoplankton and detritus; α_D and α_N represent the efficiency of the grazing; (1-α_D) and (1-α_N) the non-assimilated fractions of grazing by zooplankton that return to detritus and nutrients; rem_D is the remineralisation rate of detritus and w_D is the sinking speed of detritus.

The exact equations used differ between models, the ones given above are common examples. Other processes are often considered as well, notably respiration, excretion, and egestion, which cause loss of organic matter. Of course, additional processes may be included in more complex models.

The processes can be modelled using different mathematical forms, often with parameter values which are uncertain and can be tuned. While sinking and mortality rates are usually single parameters (linear functions), phytoplankton growth rate requires multiple parameters. μ_P is usually a function of nutrients, light and temperature:

$$\mu_P = \mu_P^{max} f(T) f(I) f(N) \quad (9.6)$$

μ_P^{max} is the maximum growth rate, $f(T)$ is the temperature effect, $f(I)$ and $f(N)$ are the limitation terms due to light and nutrients. Different formulations exist for each of these terms, but usually NPZD-type models characterise nutrient limitation of phytoplankton growth rate using Michaelis-Menten kinetics:

$$f(N) = \frac{N}{K+N} \quad (9.7)$$

K is known as the half-saturation constant for nutrient uptake, and N is the nutrient concentration. If nutrient is plentifully available, then $N/(K+N) \approx 1$ and phytoplankton growth is not limited by the nutrient.

The state variables of NPZD models represent concentrations of a given chemical element, often nitrogen, with other elements such as carbon derived using constant stoichiometry between carbon, nitrogen and phosphorus, i.e. the Redfield ratio of 106:16:1 (Redfield, 1934).

More complex models include additional variables for each compartment. Phytoplankton can be split into PFTs, grouping together species which perform a similar function within the ecosystem (Le Quéré et al., 2005). PFTs are often based on organism size. It is also common to separate out diatoms, which form silicate shells and play an important role in the sinking of carbon. In models, PFTs are distinguished by differing parameters for traits such as maximum phytoplankton growth rates, grazing, and nutrient affinity. Zooplankton can also be split into functional types, again often based on size, with different feeding preferences. Note that the current paradigm neglects the fact that many plankton are mixotrophs: they both photosynthesize and eat other organisms (Flynn et al., 2013; Glibert et al., 2019).

Variable stoichiometry (elemental ratios) can also be introduced. Each PFT is then described by separate state variables for each element, such as nitrogen, carbon, and phosphorus.

Chla is often included into BGC models as it is the main photosynthetic pigment found in phytoplankton, and measurement of its concentration in water is used as an indicator of the phytoplankton biomass. Chla can be represented as a constant ratio to the carbon biomass, or a variable ratio depending on nutrient, light levels, and temperature (Geider et al., 1997).

Most models incorporate dissolved inorganic nitrogen as a nutrient, which includes nitrate and ammonium. Phosphate and iron may be modelled too, and silicate if diatoms are a PFT. Nutrient inputs from rivers and the atmosphere can also be specified. Detritus may be split into different sizes, with different sinking rates, and into different elements. Some models explicitly simulate bacteria and viruses, rather than just parameterising their effects.

Besides NPZD variables, models can also include other related processes, such as the oxygen and carbon cycles. The carbon cycle is usually represented by the state variables DIC and total alkalinity, the latter being the capacity of seawater to neutralise an acid. From these and other variables, quantities such as pH and air-sea CO2 flux can be calculated (Zeebe and Wolf-Gladrow, 2001).

BGC models are closely related to higher trophic level models or ecosystem models. The latter require the underlying biogeochemistry, and BGC models require at least some parameterisation of the ecosystem, i.e. the explicit representation of part of the living component of the ocean (e.g. phytoplankton, zooplankton) with zooplankton mortality as a closure term, parameterising the predation of zooplankton by higher trophic levels such as fish and top predators (see Section 9.2.8).

Adding complexity to BGC models means that less important processes are neglected or amalgamated, but also increases the uncertainties associated with approximated formulations. There is no consensus on optimal structure and complexity, which will vary depending on the purpose (Fulton et al., 2003). Adding extra variables also increases computational cost, split between the computation of transport (advection and diffusion) for each state variable and the computation of the non-linear functions relating the state variables of the BGC model. In an operational context, the balance between model complexity and computational costs is critical and must be carefully evaluated. BGC models should be as simple as possible and as complex as necessary to answer specific questions.

9.1.2.2. Model calibration

As already mentioned, biogeochemical models are based on empirical relationships to describe the dynamics of biological processes. Observational data are then essential for tuning, adjusting or revising the formulations, i.e. making the model results match the observed distributions and fluxes of inorganic and organic quantities. Model calibration can be performed "by hand", i.e. by adjusting certain parameters of the biogeochemical models until the models show a "good" fit to the observed tracer fields, or by using objective optimisation methods (Kriest et al., 2020). The resulting set of biogeochemical parameters is often closely linked to the ocean circulation, mixing, and ventilation derived from the physical model used, with its specificities and defaults.

9.1.2.3. Physical-Biogeochemical coupling

Ocean physics advects and diffuses BGC model variables, thus redistributing inorganic and organic amounts. In addition, some BGC processes depend on physical conditions such as temperature or salinity, particularly crucial for the carbon cycle. Thus, there is a very strong link between the physical conditions and the BGC, which makes the BGC models closely dependent on the physical models.

Vertical motions are particularly critical to bring nutrients from nutrient-rich deep waters into the uppermost layer that receives the sunlight needed for photosynthesis and marine life. Two critical layers together regulate phytoplankton production:

- The mixed layer is the upper layer of the ocean that interacts with the atmosphere. It is assumed to be mixed and homogeneous through convective/turbulent processes, generated by winds, surface heat fluxes, or processes modifying salinity. The deeper it is, the deeper

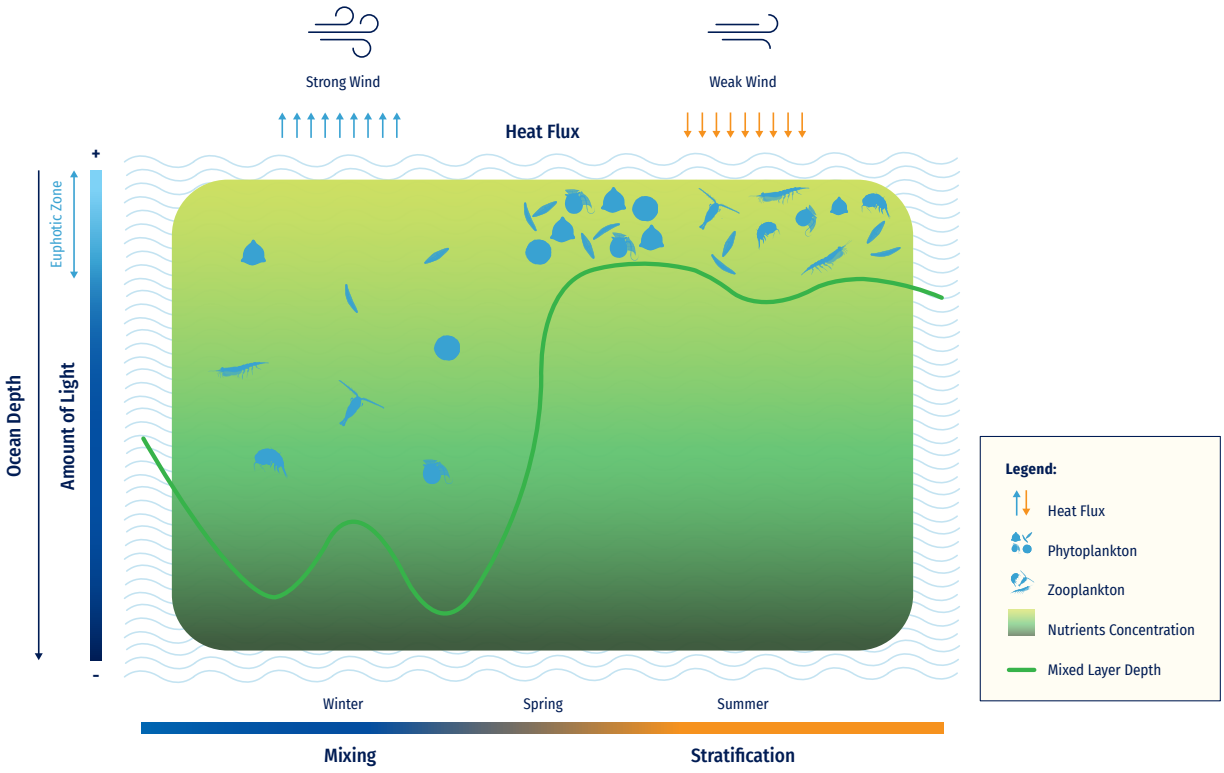


Figure 9.4. Schematic representation of the interplay between mixed layer depth (yellow line) and upper-ocean euphotic zone (light blue area) on the initiation of phytoplankton bloom (modified from Dall'Olmo et al., 2016).

phytoplankton are mixed, which will take them away from the light required for photosynthesis. Deep mixing also replenishes near-surface nutrient stocks.

- The euphotic zone is the layer from the surface down to the depth at which irradiance is 1% of the surface irradiance. The deeper the euphotic depth, the deeper the layer in which photosynthesis and phytoplankton production can occur. It extends from a few metres in turbid estuaries to approximately two hundred metres in the open ocean.

The mixed layer may develop within the euphotic layer (in stratified situations), or over a greater thickness of up to several hundred metres (in well-mixed situations). The interplay between these two critical layers controls the plankton exposure to sunlight and the coincident exposure to nutrients, thus regulating phytoplankton production (Figure 9.4). Exact mechanisms are still debated. Please refer to Ford et al. (2018) for more details.

In turn, phytoplankton abundance may feed back to physics, by absorbing radiation in the surface layers and therefore affecting heat penetration into the water column (Lengaigne et al., 2007).

9.1.2.4. From open ocean to coastal ecosystems

Different considerations are generally needed for open ocean and coastal ecosystems. In the open ocean, the seasonal cycle is quite well defined and recurring (Figure 9.5). Seasonal increases in temperature and solar radiation drive the phytoplankton spring bloom. The peak persists for a few weeks to months until nutrient limitation and grazing cause the bloom to collapse. A secondary biomass peak can develop in late summer or autumn.

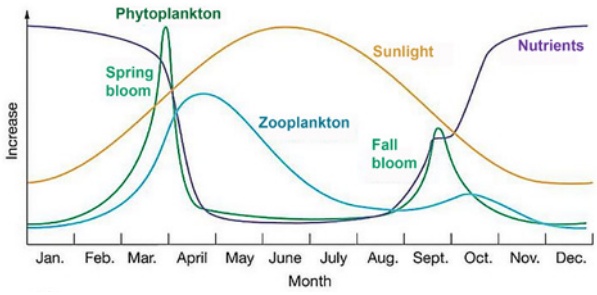


Figure 9.5. Seasonal cycle of phytoplankton relative to variations in sunlight, nutrients, and zooplankton (Copyright: 2004 Pearson Prentice Hall, Inc).

In contrast, coastal ecosystems can be very complex, subject to a succession of blooms having different origins, thus requiring additional model complexity. Correct specification of river inputs also becomes more critical. Furthermore, the equations in Section 9.1.2.1 are for the pelagic (water column) ecosystem. In shallow waters, such as shelf seas, it becomes important to include the benthic (seafloor) ecosystem into the BGC models. This requires the addition of extra variables, though they do not need to be advected or diffused. Finally, coastal waters are often turbid, and the effect of sediments and coloured dissolved organic matter on light and therefore primary production should be included. Dedicated optical models are sometimes used for this purpose (Gregg and Rousseaux, 2016).

9.1.2.5. Potential predictability of ocean biogeochemistry

The potential predictability of ocean biogeochemistry varies considerably depending on the scales and quantities of interest. A lot of variability is driven by physics, with changes in mixing and stratification affecting light and nutrients and therefore primary production. When these physics changes can be predicted, e.g. changes in stratification with a warming climate and interannual variability related to phenomena such as the El Niño Southern Oscillation, associated large-scale changes to ocean biogeochemistry can also be predicted. Similarly, changes to the ocean carbon cycle and acidification with increasing atmospheric CO2 concentrations can be predicted. When considering local regions and/or shorter time scales, both physics and biogeochemistry become harder to be accurately predicted.

Furthermore, various biogeochemical quantities change at very different rates. Phytoplankton react quickly to changes in light and nutrient availability and can double in concentration over a day (Laws, 2013). Zooplankton will exhibit a slightly more lagged response to these changes. Meanwhile, nutrient concentrations will typically change more slowly, and the carbon cycle even more slowly, although surface concentrations (of nutrients and carbon) can change rapidly, for example during a storm. These different rates of change have implications for the scales of predictability.

For accurate predictions, it is important to initialise models using data assimilation (see Section 9.2.5). At seasonal-to-decadal time scales, predictability is dominated by physics, and this must be accurately initialised and simulated. Physics remains important at shorter time scales, but is essential to initialise nutrient concentrations correctly, as this will help to determine the primary productivity. For short-range predictions, phytoplankton concentrations should be initialised, though the memory of the phytoplankton variables may be as short as a few days, given that they react to changes in nutrients and mixing. Accurate model formulations and parameterisations are also required, otherwise the model will react incorrectly to the data assimilation.

9.2. Biogeochemical forecast and multi-year systems

Green Ocean modelling for operational oceanography is built in the same way as its Blue equivalent. The operational suite follows almost the same architecture (see Figure 4.1) and information flows from marine observation data up to end-user products enhancing the initial information. Each component includes a research stage, a development stage, and an operational stage. This Chapter mainly focuses on the last stage, in which the system is in operation.

The modelling component includes the BGC model, data assimilation, and ensemble modelling, executed for analysis and to forecast BGC conditions. The data include upstream data such as physical conditions, atmospheric forcing, external inputs of chemical compounds provided at interfaces (atmosphere, land, and seafloor), observational data from satellites, and in-situ measurements integrated into the systems via data assimilation methods. The data are also used for validation tasks: the near-real time evaluation of the forecast accuracy and the delay mode evaluation of the model system. Finally, the model outputs and end-user products are prepared by respecting certain standards of format, units, names, etc. for delivery to users and archiving.

9.2.1. Architecture singularities

In this section, we present the main architecture singularities of OOFS dedicated to the production of ocean biogeochemistry and marine ecosystems information. As most systems describing the "Green Ocean" in operation today are less advanced than their "Blue Ocean" equivalent, the "ideal" design proposed here includes some features that are still at the stage of research or development. Yet, they should be kept in mind for the construction of future systems.

9.2.1.1. Physical, optical, and biogeochemical components

As introduced in Section 9.1.2, the space-time evolution of the BGC quantities is driven by physical fields through horizontal and vertical advection, lateral diffusion, and vertical mixing. Vertical motions are particularly important as they supply nutrients to the lighted upper ocean, allowing photosynthesis to occur.

The limitation of photosynthesis by light thus requires a fine representation of the penetration of spectral irradiance in the upper ocean, as it is absorbed and scattered within the water column. Light penetration used to be managed by very simple optical schemes, but it is now increasingly managed by advanced bio-optical modules embedded into the physical-biogeochemical model systems, to both compute photosynthetic activity and to make the link with key observations such as spectral irradiances from ocean colour missions. The evolution of ecosystem variables in the trophic chain is driven by physics, optics, and biogeochemistry through primary production, which underpins the whole marine ecosystem (see Section 9.2.8).

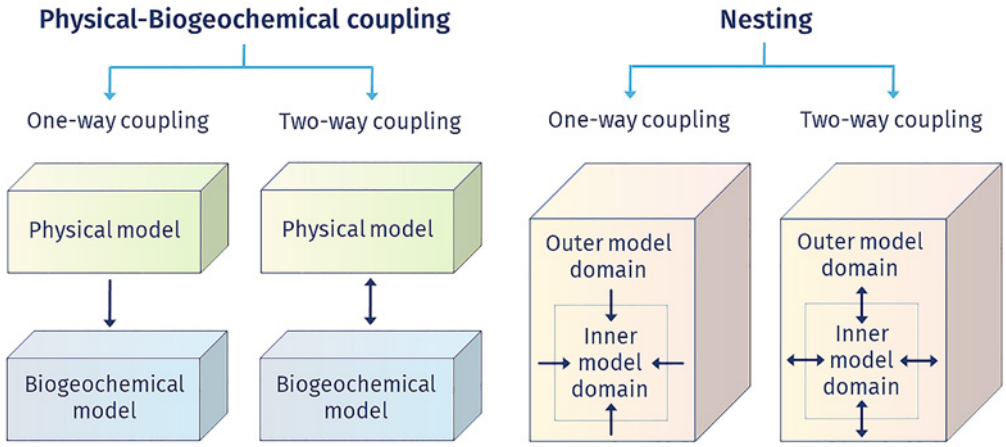


Figure 9.6. Schematic of a physical-biogeochemical coupling (left) and nesting (right).

The physical fields can come from simulations of the ocean dynamics (reanalyses, nowcasts or forecasts) produced independently by the BGC modelling suite. Nevertheless, the physical fields must reflect the essential dynamical properties for the biogeochemistry, such as the right mixing rate, the right vertical velocity statistics, and the right phasing with mesoscale structures and frontal positions (Berline et al., 2007).

Although some feedback from biology to physics may exist, such as self-shading (Hernandez et al., 2017) or phytoplankton heat release, their effects are generally limited to moderate modifications of the upper-ocean heat budget and associated vertical structure of the thermocline. Therefore, the physical and BGC modelling components are usually linked by "one-way" coupling, resulting in successive model operations (Figure 9.6). As a result of the "one-way" approximation, the coupling can be implemented in "online" mode, i.e. the physical and biogeochemical models run simultaneously at each time step: the temporal update of the physical model is performed first, before being used for the update of the biogeochemical component. Alternatively, the coupling can be implemented in "offline" mode where the physics is computed beforehand and stored at lower frequency (e.g. each day/week) and then used as inputs for the biological model (Ford et al., 2018).

Such systems are usually less expensive in terms of computational resources. However, the practicality of the "offline" coupling approach can be questioned with respect to vertical viscosity and diffusivity coefficients, which typically vary with short time scales (~hours) compared to the storage rate of "offline" physical fields (typically a few days). This can be an issue in an integrated perspective that includes data assimilation. Burning questions underlying the coupling strategy for assimilative systems are still the subject of long-lasting research efforts by the community (Fennel et al., 2019).

Regional models with lateral open boundaries also require values of the model state variables at boundaries. A convenient way is nudging to fixed or climatological data from global reanalysis or datasets, but a more robust approach is to nest high-resolution regional ocean models into larger-domain (and usually lower-resolution) models (see Figure 9.6).

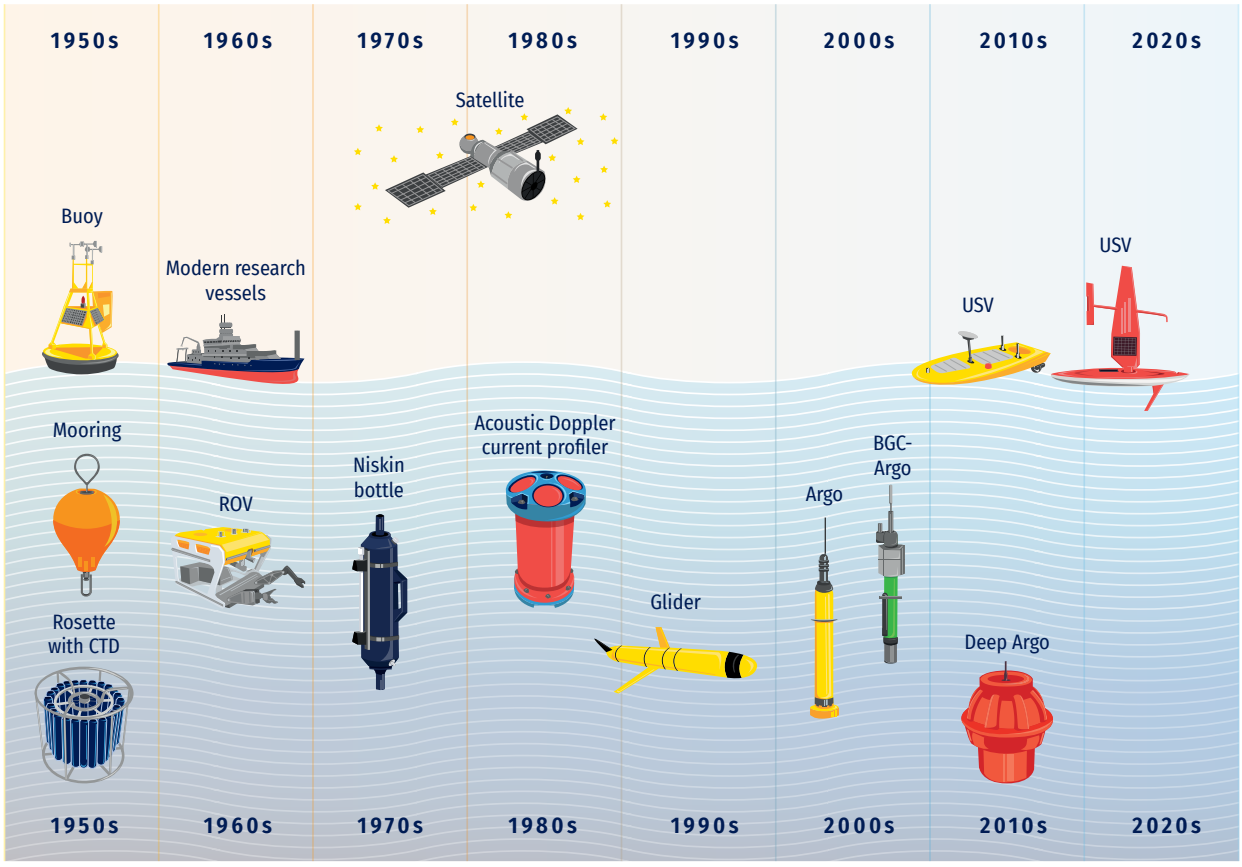


Figure 9.7. Chronology of oceanographic observation platforms to measure marine biogeochemistry (adapted from Chai et al., 2020).

As for the coupling between physics and biogeochemistry, the coupling between configurations nested in space can be "one-way", with the inner model having no influence on the outer model, or "two-way", in which the inner model provides information to the outer model. "One-way" coupling is mainly used in BGC operational systems for different reasons, as it offers the possibility to run the BGC model either in "online" or "offline" mode with the physics, while the "two-way" nesting requires by nature an "online" coupling between the physics and the BGC, making the operation of such coupled systems more complex and time-consuming.

For a sound representation of the biology, a specific design of the vertical discretization in the upper ocean is needed. The strong vertical gradients of the physical and biological variables typically require vertical spacing between horizontal levels ~ 1 metre. Regarding the horizontal grid, it is not always required to use the same numerical grid for physics and for biology. A coarsening approach that preserves the essential features of the resolved dynamics has been implemented in some systems to feed the biological equations at lower resolution, while saving numerical resources (Berthet et al., 2019; Bricaud et al., 2020).

9.2.1.2. Propagation of uncertainties

The forward integration of the discretized equations involved in the different modelling steps leads to results that are fundamentally uncertain. It is necessary to quantify this uncertainty, both to provide the user with useful information for decision making and for merging the forecast with future observations, which are also intrinsically uncertain.

The main possible sources of uncertainty in biogeochemical/ ecosystem models are the following:

- initial conditions of the state variables;
- external data involved in the forcings, such as downward radiation, cloud cover, etc.;
- input physical data used to constrain the evolution equation of the biogeochemical/ ecosystem variables, such as currents, temperature, vertical eddy viscosity, etc.
- parameters involved in the representation of optical, BGC and ecosystem processes;
- numerical schemes and numerical approximations (such as coarsening or offline integration);
- unresolved, sub-grid scale processes that may induce bulk effects as a result of non-linearities.

These uncertainties can be quantified heuristically or can be explicitly considered by introducing stochastic parameterizations in the model equations, as proposed by Garnier et al. (2016). Multiple forward integrations can then be produced to generate ensembles that provide an approximation of the spread of the plausible solutions. A sample of the prior probability distribution of the forecast is then generated by the different ensemble members (Santana-Falcon et al., 2020). As a result, the forward integration module (referred as Step 2 "Forecast" in Figure 4.1) should be designed in such a way that it can be called n times (with n = a few tens to hundreds) in parallel or in sequence. Please refer to Section 9.2.4 for more details on Ensemble modelling.

9.2.1.3. BGC Data singularities

Biogeochemical variables very often have non-Gaussian statistical properties. This can be explained by the nature of these variables (generally concentrations that repeatedly take values close to 0 or biomasses that can vary by several orders of magnitude), which is related to the non-linearities of the processes involved. Non-Gaussian behaviour requires special attention at the time of validation when comparing model variables to observations, using metrics calculated on log-transformed data or non-parametric metrics (please refer to Section 9.2.6 for more details).

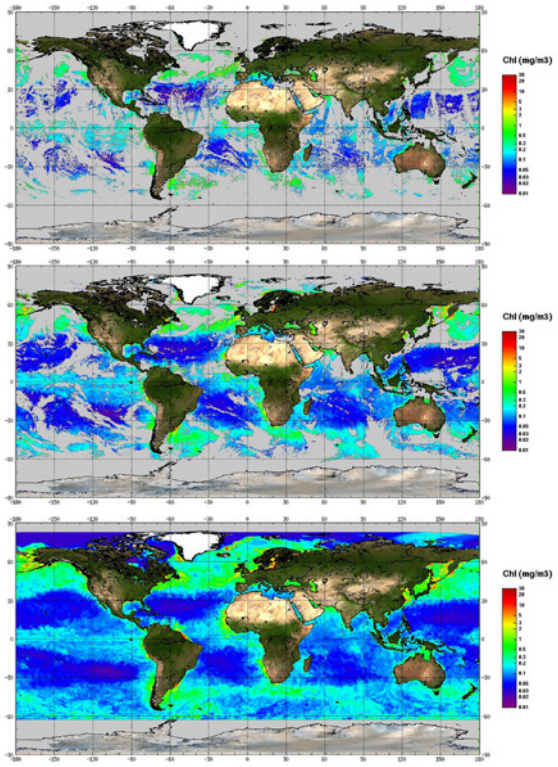


Figure 9.8. Examples of Chla ocean colour global multi sensor products available on the Copernicus Marine Service. They are daily products for 1st May 2019: a) OC-CCI product; b) Copernicus-GlobColour level 3 product; and c) Copernicus-GlobColour "Cloud Free" (interpolated) product (from Garnesson et al., 2021).

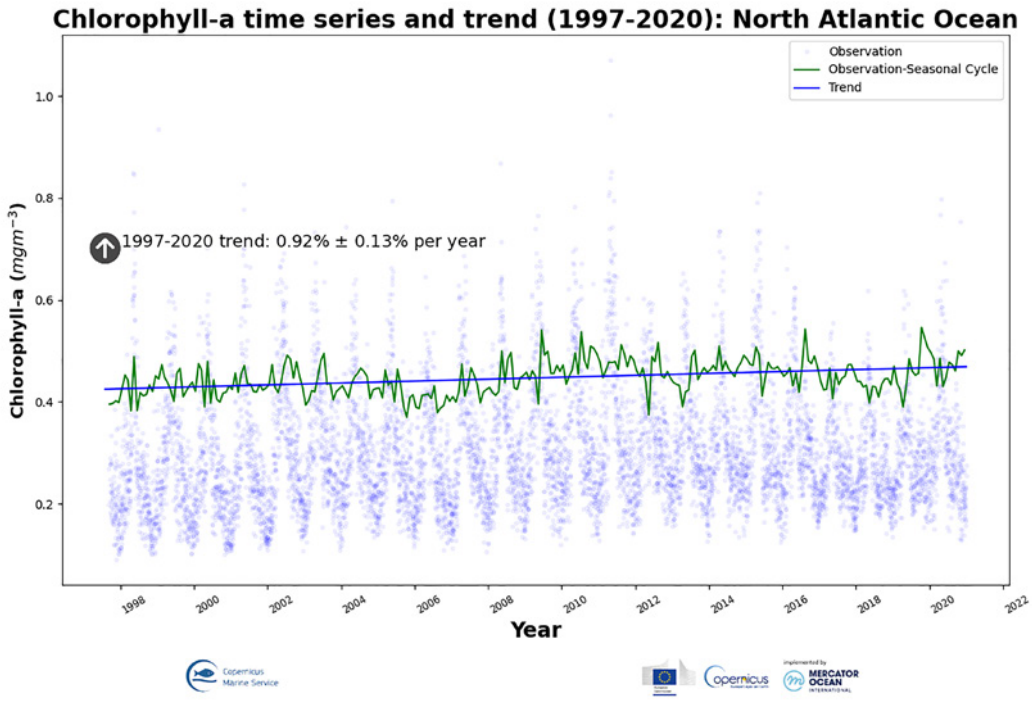


Figure 9.9. North Atlantic Ocean time series and trend (1997-2019) of satellite chlorophyll. Blue dots: daily regional average time series; green line: deseasonalized time series; blue line: linear trend (source: Copernicus Marine Service at [2]).

In addition, the assimilation methods applicable to large systems, e.g. Ensemble Kalman Filters, are typically adapted to Gaussian distributions: as a result, it is necessary to insert a so-called anamorphic transformation – a function matching the quantiles of the variable distribution to those of a standard Gaussian – between the outputs of the ensemble forward integration and the observational update step. This can be done in different ways: by prescribing a priori a given transformation (e.g. log-normal or truncated Gaussian), or by constructing the transformation from the ensemble information as proposed by Simon and Bertino (2009 and 2012) and Brankart et al. (2012). At the end of the analysis step, the inverse transformation must be applied to complete the assimilation cycle and prepare a new initialization.

Another issue comes from the highly heterogeneous distribution of the biogeochemical data in space and time, most of which coming from satellites (ocean colour) and fairly dispersed BGC-Argo profilers. The spatial scales captured by these observational data are therefore very different, requiring special care within biogeochemical data assimilation systems for localization at the analysis stage. The transformation in the Fourier space can then prove beneficial to carry out this step, as proposed by Tissier et al. (2019). The architecture of an operational chain dedicated to biogeochemistry should therefore include a step to perform the observational update in a transformed space.

9.2.2. Input data: available sources and data handling

This Section provides a general description and technical information on the data used to both drive and validate a biogeochemical forecasting system. Observational data are required at different stages of an OOFS:

- Data is first used to set-up the model configuration: initial and lateral conditions, physical forcing, atmospheric surface forcing, and external inputs.
- Data is essential for calibrating the formulations of the BGC processes, i.e. making the model results to match the observed distributions and fluxes.
- Then data is used to evaluate the model product quality.
- Finally, observational information is incorporated into the numerical models using data assimilation methods with the objective to improve predicted model states.

2. https://marine.copernicus.eu/access-data/ocean-monitoring-indicators/north-atlantic-ocean-chlorophyll-time-series-and-trend

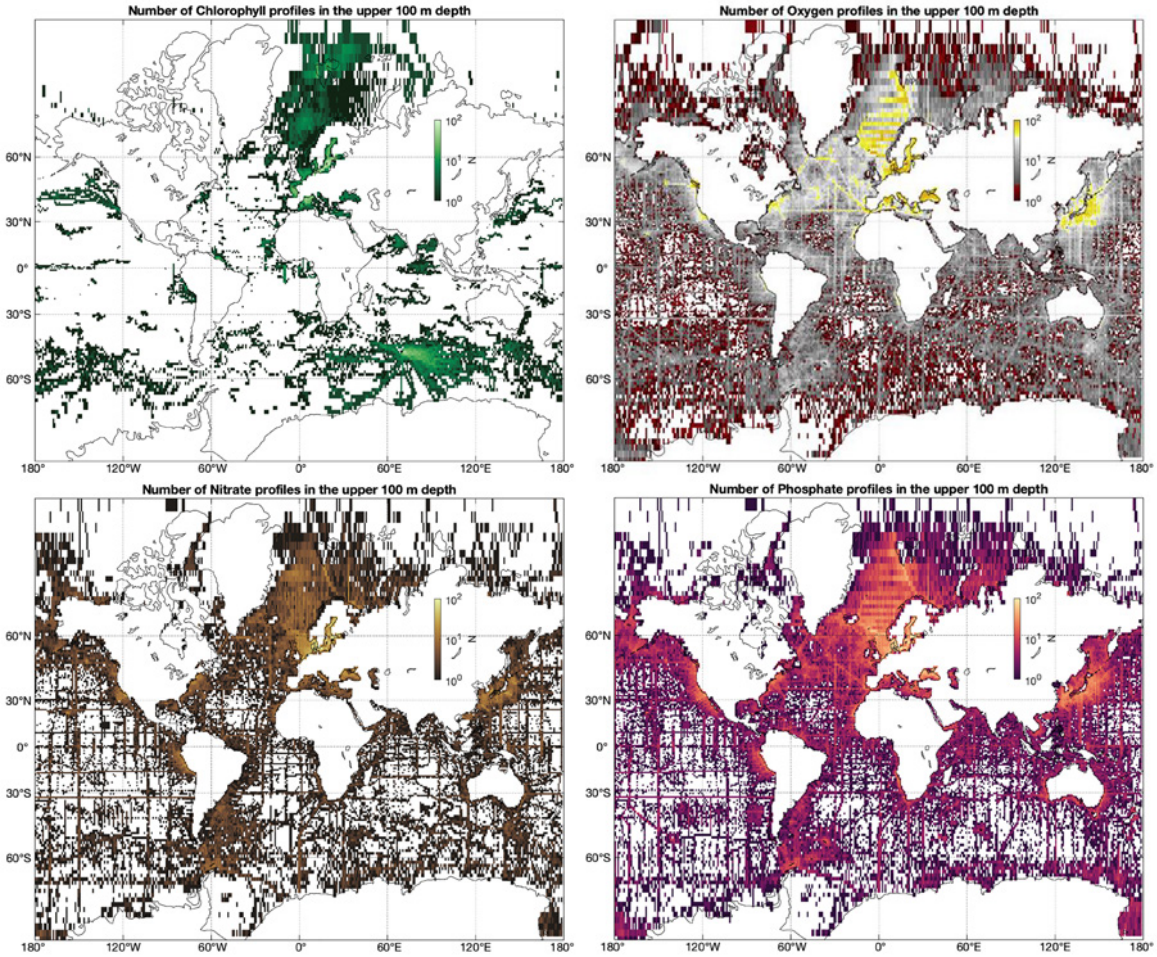


Figure 9.10. Spatial coverage of chlorophyll (top left), oxygen (top right), nitrate (bottom left) and phosphate (bottom right), shown as the number (N) of profiles in the upper 100 m water depth in 1°x1° cells, from 1990 to 2020. To show gaps more clearly, colour shading is from dark (low sampling) to light (high sampling), white colour indicates no sampling (from Jaccard et al., 2021).

9.2.2.1. Physical conditions

Required fields are currents, temperature, salinity, vertical diffusivity coefficient (Kz), and MLD. They are provided by a physical model to the BGC model with which it is coupled in either "online" or "offline" mode (see Section 9.2.1 for details). Advection and diffusion routines are usually shared with the physical model. A list of physical-BGC coupled systems is available in Section 9.2.9.

9.2.2.2. Observational data

Ocean-observing platforms to measure marine BGC encompass ship, mooring, and remote sensing observations. A good overview of the evolution and diversification of platforms over the past century is given by Chai et al. (2020) from which is taken Figure 9.7. Among the traditional observing systems, satellites represented a revolution, providing a continuous spatiotemporal coverage of sea surface variables. More recently, autonomous mobile platforms measure ocean variables through the water column. They cover a wide range of spatial and temporal scales, filling the observational gaps.

9.2.2.2.1. Remote sensing observations

Remote sensing-derived Chla data have a good spatial coverage of the entire ocean in near-real time and reprocessed time series for global and regional mapped products. They are available through operational services, such as the Copernicus Marine Service (🔗[3]; Le Traon et al., 2017). Figure 9.8

3. https://marine.copernicus.eu/

presents some Chla products and their spatial coverage. Figure 9.9 illustrates the long time series available. Remote sensing derived PFTs and optical properties are also starting to be distributed on the same portal.

9.2.2.2.2. In-situ observations

The Copernicus Marine Service collects and distributes in-situ observations from a variety of platforms, including manual CTD-O2 measurements, BGC-Argo profiling floats, ferrybox systems, gliders and moored buoys, gathered by global systems such as the EuroGOOS, SeaDataNet, NODCs, and the JCOMM. Two types of products are provided: 1) NRT products automatically quality controlled within 24 hours from acquisition for forecasting activities and 2) the reprocessed (or multi-year) products for reanalysis activities. The main biogeochemical variables available are dissolved oxygen concentration, nutrients (nitrate, silicate and phosphate), Chla, fluorescence, and pH. The spatial distribution of all chlorophyll, oxygen, nitrate and phosphate samples of the reprocessed product (from 1990 to 2020) are shown in. Figure 9.10.

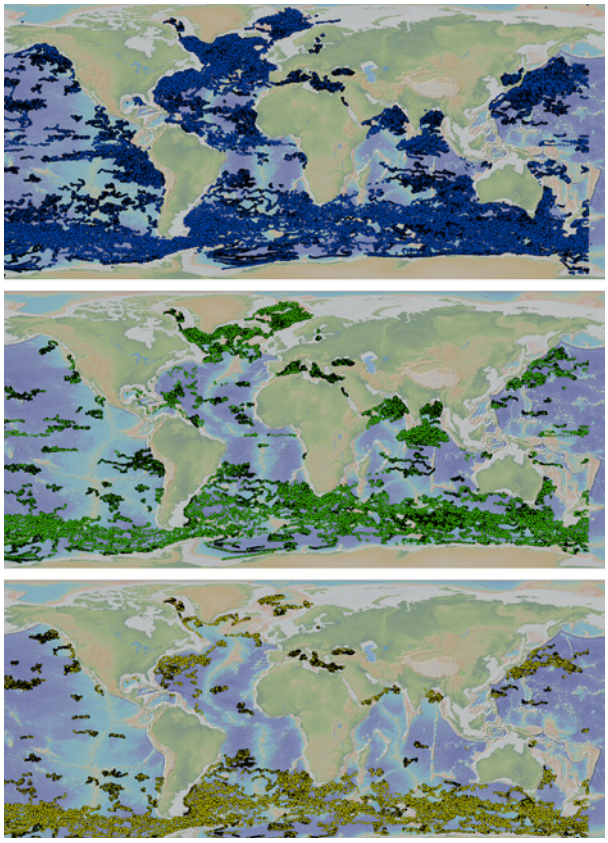


Figure 9.11. Spatial coverage of oxygen (top), Chla (middle), and nitrate (bottom) from the start of the BGC-Argo program. 230,202 profiles of oxygen, 94,947 profiles of Chla, and 49,939 profiles of nitrate have been acquired by October 2021 (source: T. Carval, personal communication using data from the Copernicus Marine Service).

Special attention should be paid to autonomous robotic underwater vehicles. Argo profiling floats drift freely with the currents and measure ocean variables through the water column, reaching up to 2000 m, while gliders can be programmed to sample along a predetermined path, making the former more suited to the open ocean and the latter more suitable for observation at various depths in coastal and shallow oceans. After cycling vertically, both floats and gliders transmit their data to orbiting satellites once they have reached the surface, providing continuous monitoring and real-time data to operational centres.

The International Biogeochemical-Argo (BGC-Argo) program is revolutionising marine biogeochemistry by establishing a global, full-depth, and multidisciplinary ocean observation network, acquiring profiles in regions of the global ocean that previously were observationally sparse (Russell et al., 2014). They measure oxygen, Chla, nitrate, pH, suspended particles, and downwelling irradiance. Since their deployment in 2012, 1623 floats have acquired about 250000 profiles (Figure 9.11), the major part being oxygen. The aim is to have 1000 active profiling floats measuring simultaneously the six essential variables mentioned above (Biogeochemical-Argo Planning Group, 2016; Chai et al., 2020). At the time being, 410 floats are operational around the world (Figure 9.12). An example of time series is presented in Figure 9.13. BGC-Argo data are publically available in near real-time after an automated quality control, and in scientifically quality controlled form, delayed mode data, within six months of collection, via two Global Data Assembly Centers (Coriolis in France and US-GODAE in USA) (Argo, 2022; [4]). They are also available through the Copernicus Marine Service ([5]).

4. https://www.seanoe.org/data/00311/42182/
5. https://marine.copernicus.eu/

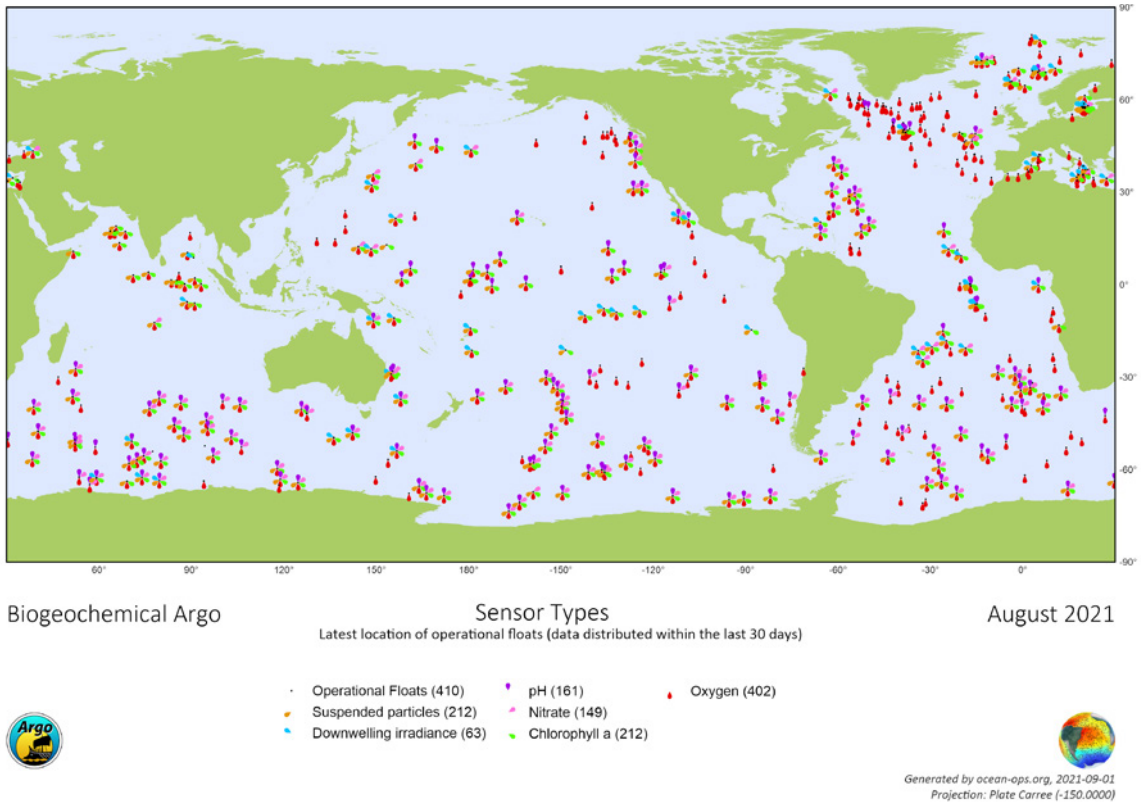


Figure 9.12. Location of operational BGC-Argo floats in August 2021 (🔗[6]).

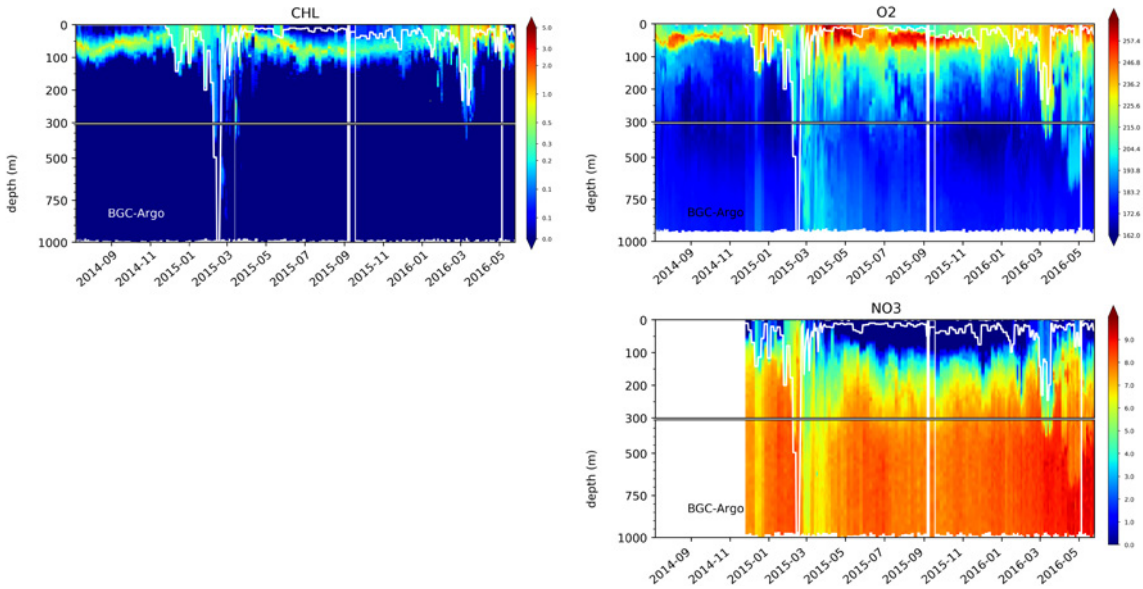


Figure 9.13. Time evolution of Chla (top left), oxygen (top right) and nitrate (bottom) along a BGC-Argo float trajectory in the North-East Atlantic.

6. www.ocean-ops.org

9.2.2.3.Climatologies, databases, and atlases

Databases and atlases are collections of uniformly formatted, quality controlled, and publicly available ocean surface or vertical profile data. Climatologies are mapped data products, produced from databases and atlases, representing the mean annual, seasonal, or monthly large-scale characteristics of the distribution of a quantity. They can be used to create initial and/or boundary conditions for ocean BGC models, evaluate numerical simulations, and corroborate satellite data.

The GLODAP provides a climatology (GLODAPv2.2020) of ocean biogeochemical variables of oxygen, phosphate, nitrate, silicate, dissolved inorganic carbon, total alkalinity, and pH on a uniform 1° longitude/latitude grid. The product is described in Olsen et al. (2020) and is publicly available at [7].

The latest version of the WOA delivered in 2018 provides an annual, seasonal, and monthly climatology of oxygen and macronutrients (phosphate, silicate, and nitrate) on a 1° longitude/latitude grid (Figure 9.14).

It is described in Garcia et al. (2018ab) and is publicly available at [8]. It is based on the latest major release of the WOD described in Boyer et al. (2018).

The SOCAT provides surface ocean fCO_2 (fugacity of carbon dioxide) observations, [9]. The latest SOCAT (version 2020) has 28.2 million observations from 1957 to 2020 for the global oceans and coastal seas.

The EMODnet portal provides access to temporal and spatial distribution of marine chemistry data in European seas, [10].

9.2.2.4.Atmospheric surface forcing

Atmospheric surface conditions drive biogeochemical quantities and processes, such as photosynthesis and air-sea exchanges of gas elements (oxygen, carbon). Typical surface data inputs include wind, solar radiation, and the evaporation-precipitation flux. They can be obtained from an operational weather prediction system, via the Copernicus Climate Change Service ([11]).

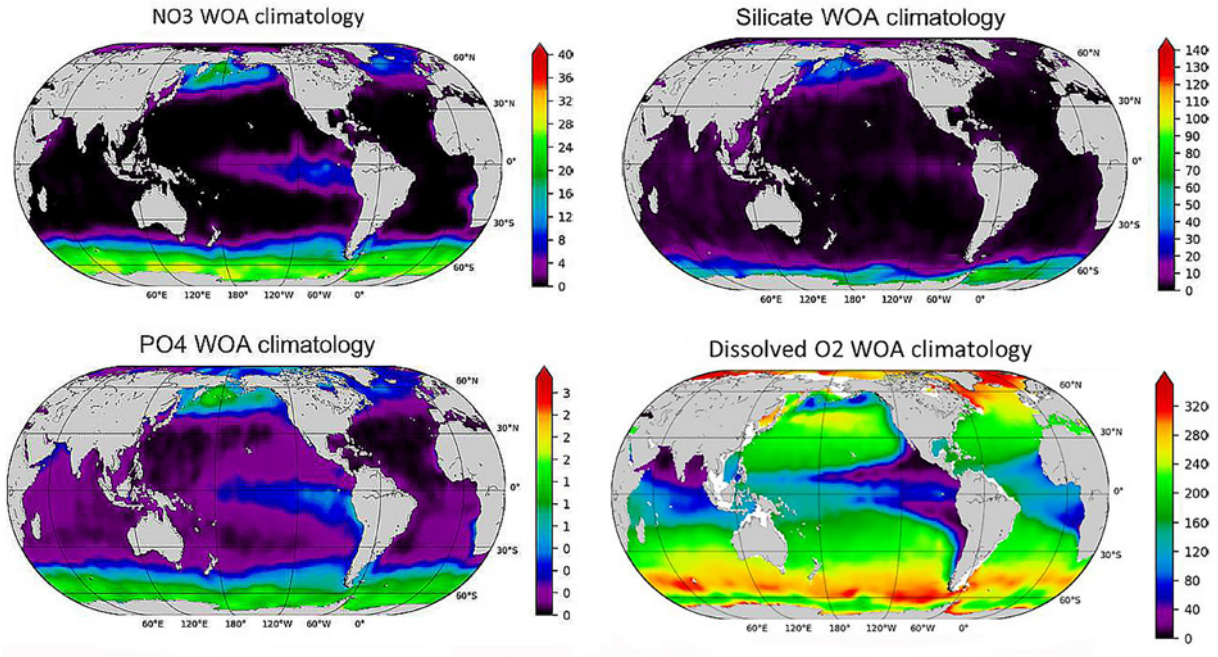


Figure 9.14. Nitrate, phosphate, and silicate concentrations at sea surface and dissolved oxygen concentration at 200 m depth, all in mmol m^{-3} (from WOA climatology).

7. https://www.glodap.info
8. https://www.ncei.noaa.gov/products/world-ocean-atlas
9. https://www.socat.info/
10. https://emodnet.eu/en/chemistry
11. https://climate.copernicus.eu/

9.2.2.5. External inputs

External inputs of carbon and nutrients are provided to marine biogeochemical systems from observations or models. Although these inputs are currently simplified in current systems (from climatologies), the optimal solution would be to connect ocean operational systems with atmospheric and land operational systems. The link between the Copernicus Marine Service and the Copernicus Atmosphere and Land Services (respectively, [12] and [13]) is currently discussed.

9.2.2.6. Units

Special attention should be paid to the units of the BGC quantities because there is no standardisation among the different scientific communities. Model data are usually archived in the units specified by the SI Units but instruments frequently do not measure data in SI Units, making conversion necessary. For example, dissolved oxygen concentration in the seawater can be found in many different units (e.g. mg l^{-1}, ml l^{-1}, µmol l^{-1}, µmol kg^{-1}, mmol m^{-3}, µM), with the SI Units being mole per cubic metre (symbol mol m^{-3}).

It is worth noting the equivalences:

$$\mu mol\ l^{-1} = mmol\ m^{-3} = \mu M$$

$$1\ l = 10^{-3}\ m^3 \approx 1.025\ kg$$

and the conversions:

$$\mu g\ l^{-1} = \mu mol\ l^{-1} \times MW$$

$$\mu l\ l^{-1} = \mu mol\ l^{-1} \times MV$$

$$g\ l^{-1} \approx g\ kg^{-1} \times 1.025$$

To convert a quantity in sea water from mole concentration (in mol) to mass (in grams), multiply by Molar weight (MW in g mol^{-1}); from mole concentration (in mol) to volume fraction (in litre), multiply by Molar volume (MV in l mol^{-1}); expressed per unit mass (in gram) to volume (in litre), multiply by density (in kg l^{-1}). 1.025 is an approximate but general value for the density of seawater.

9.2.3. Modelling component

9.2.3.1. Numerical and discretisation choices

Marine biogeochemical models describe the cycling of essential elements (e.g. C, N, O2, P, and Si) through the lower trophic levels, usually from bacteria up to mesozooplankton. Their complexity (i.e. number of state variables and processes) differs depending on the scientific question under interest, the information available for their parameterization and implementation, and the investigated time and space scales. BGC models consist of a set of evolution equations (e.g. differential equations) expressing the mass balance of each model component (e.g. state variable). These mass balance equations include local sources and sinks associated with biogeochemical processes (e.g. photosynthesis, respiration, and nitrification), trophic interactions (e.g. predation), the transport by physical processes in the three directions of space by advection (e.g. transport by the main current), and diffusion (i.e. unresolved processes that are parameterized on the model of the Fick's law of diffusion). As for physical models, biogeochemical models cannot be solved analytically and require a numerical model for their integration. A numerical grid has to be defined and the size of the grid cells will define the spatial scales that can be solved (it is usually assumed that the length scale of the solved processes equals twice the size of the grid). Given that the vertical scales of variations are much smaller than the horizontal ones due to the rapid extinction of the light field, the size of the vertical mesh is usually of the order of metres in the upper layer. The numerical scheme for time steps and time integration has to be carefully chosen in order to avoid generating negative concentrations. The choices may be identical to the physical model to which it is coupled, or different. Numerical and discretization techniques are described in Chapter 5 and biogeochemical singularities are discussed in Section 9.2.1.

Whether the processes can be resolved or not in models will depend on the grid resolution used to solve the numeric. Figure 9.15 shows the spatial and temporal scale of specific biogeochemical processes.

Regional and global scale models are able to capture the mesoscale signals with temporal scales of the order of a month and spatial scales of the order of 50-100 km. Coastal models have to solve the high frequency signal at daily and (sub)-mesoscale, but at this stage they are able to solve the dynamics of the system at weekly to monthly scales.

9.2.3.2.The different biogeochemical models

In marine biogeochemistry, the specificity lies mainly in the diversity of environments, ecosystems, and processes. The choice of a BGC model will thus depend on the study area and the topic of interest.

Models of marine biogeochemistry and of the lower trophic levels in the marine food web are usually of the NPZD type (see Section 9.1.2 for more details), which resolve community structure by the explicit representation of a few plankton groups, in accordance with their function in the ecosystem. Another approach is to let the community structure emerge from a wide

12. https://atmosphere.copernicus.eu/
13. https://land.copernicus.eu/

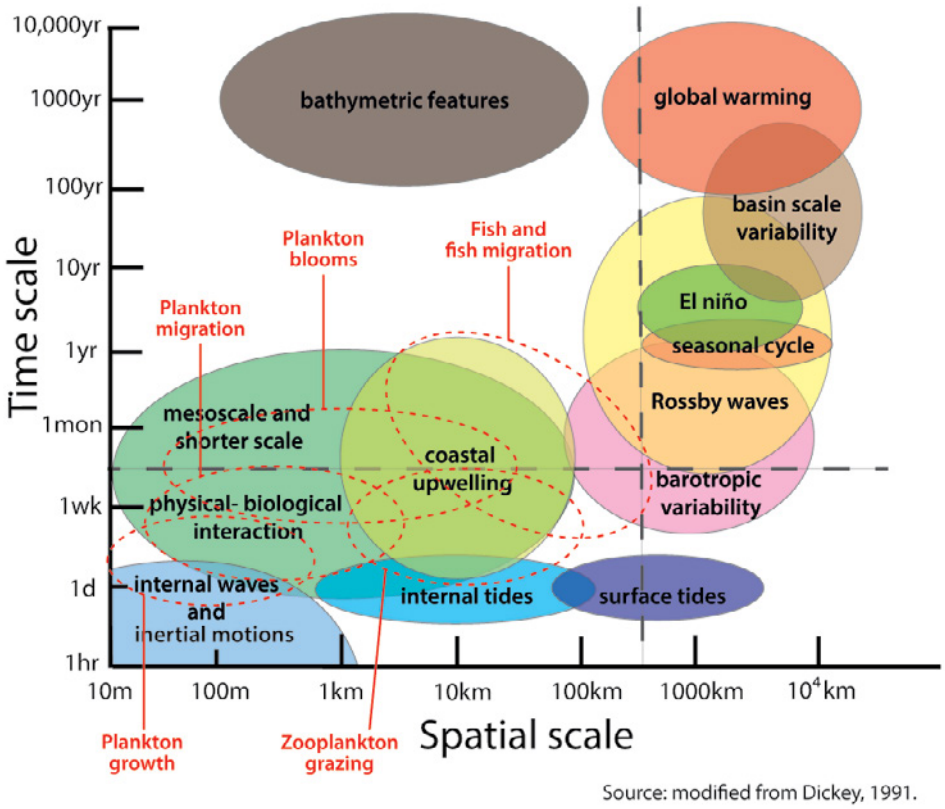


Figure 9.15. Time and space overlapping scales of major ocean processes. Main processes modelled by biogeochemical models are outlined in red (adapted from Dickey, 1991).

range of possibilities. For example, the DARWIN model (Follows et al., 2007) includes a large number (tens or hundreds) of PFTs whose physiological characteristics are stochastically determined (the parameters are prescribed randomly), allowing the fittest to emerge in the resulting ecosystem.

Some of the most used models in OOFS are summarised below:

- HadOCC (Palmer and Totterdell, 2001).
- MEDUSA (Yool et al., 2013).
- PISCES (Aumont et al., 2015). Its development is led by the Pisces Community gathering eight international research institutes/laboratories. The model can be downloaded from the NEMO and CROCO modelling systems into which it is embedded ([14] and [15]).
- ERSEM (Baretta et al., 1995; Butenschön et al., 2016). Its development is led by the Plymouth Marine Laboratory and the code is available at [16].
- BFM (Vichi et al., 2015). Its development is led by a consortium of five members and the code is available at [17].
- NORWECOM (Skogen, 1993; Skogen and Søiland, 1998). NORWECOM is the result of the cooperation between several Norwegian institutions, for more information see http://www.ii.uib.no/~morten/norwecom.html.
- ECOSMO (Daewel and Schrum, 2013) is developed by Hereon with contributions from the Nansen Centre and other collaborators, see [18].
- ERGOM (Neumann, 2000). It was developed at IOW, Germany.
- BAMHBI (Grégoire et al., 2008; Grégoire and Soetaert, 2010; Capet et al., 2016).
- SCOBI, described in Eilola et al. (2009) and Almroth-Rosell et al. (2015).

Usually, these models are the result of the collaboration between different national and international research/academic institutes and laboratories, organised in formal or informal consortia. They are shared by several operators. In most cases, the code is available under open-source licences.

14. http://www.nemo-ocean.eu
15. https://www.croco-ocean.org
16. https://www.pml.ac.uk/Modelling/Home
17. https://bfm-community.github.io/www.bfm-community.eu/
18. https://www.hereon.de/institutes/coastal_systems_analysis_modeling/matter_transport_ecosystem_dynamics/models/index.php.en

Models have been developed to be applied to regional, shelf-sea, basin, or global ocean scale. The level of complexity differs depending on the application (biogeochemical cycling or ecological application). The models mainly differ in the biogeochemical cycles of major elements, the number of nutrients, the number of autotrophic and heterotrophic PFTs, the complexity in process formulation, as well as in the consideration of the benthic component. See refer to Gehlen et al. (2015) for a detailed description of these models.

The practical ability to switch between different physical and biogeochemical models is desirable to compare models and upgrade them smoothly. This ability is offered by the FABM (🔗[19]) and it has been used in NEMO and HYCOM, among other ocean/lake models programmed in Fortran.

9.2.3.3. Connections Ocean-Earth systems

Several kinds of models are used for a range of environments, but different considerations are needed for open ocean, regional, and coastal ocean. Moving from the open to the coastal ocean is often accompanied by an increase in the spatial resolution and complexity of the model.

Regional models of coastal ecosystems can be very complex. Their dynamics is essentially driven by the boundary conditions with the open sea and at the air-sediment-land interface.

For the ocean, atmosphere, rivers, and sediments are significant sources of carbon and bioactive nutrients, such as nitrogen, phosphorus, iron, and silicate. Model performances can be hampered by the quality of these boundary conditions. Coastline and topography are also important to trigger high-frequency physical processes.

Connections with the surrounding systems (Figure 9.16) that need to be carefully considered include:

- Connection with land. Rivers exchange freshwater as well as inorganic and organic material with the ocean. Coastal marine ecosystems have been subject to considerable modification in recent decades. The considerable nutrient load in river discharges is due to human activities on the land (e.g. agriculture, deforestation, waste discharge, etc.). Such inputs are critical for coastal ecosystem studies.

- Connection with the atmosphere. Atmospheric transport and deposition are a source of chemical compounds (e.g. carbon dioxide, nitrogen, oxygen, iron, and phosphorus) to the ocean, affecting marine biogeochemistry (e.g. source of nutrients, influence on pH, etc.) (Krishnamurthy et al., 2010).

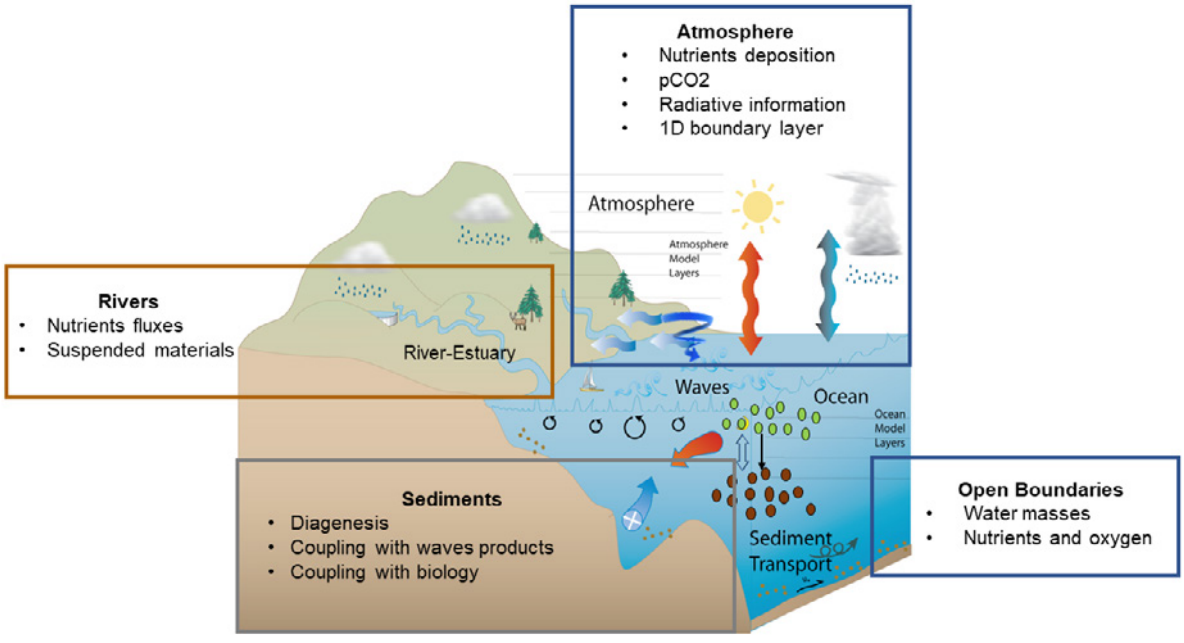


Figure 9.16. Connections with interfaces (modified from Warner et al., 2010).

19. https://bolding-bruggeman.com/portfolio/fabm/

- Connection with the seafloor. Exchanges between the sediments and the ocean can be represented in a very basic way: they consist of the deposition of non-living organic material, resuspension, and release of inorganic nutrients from the sediments. But for a more robust approach it should be used an additional module representing (semi-) explicitly the diagenesis, benthic ecosystem, as well as bioturbation, diffusion, bio-irrigation effects into the upper sediments and sediment transport. A coupling with the waves is sometimes realised, e.g. using climatology.

- Connection with the open ocean. Open ocean and coastal ecosystems are intimately linked as they exchange mass, fluxes, and materials with each other. The best possible knowledge of open boundary conditions is essential for coastal modelling.

- The sea ice algae contribute between 4 and 26% of the primary production in the sea ice covered regions of the Arctic Ocean (Spindler, 1994; Gradinger, 2009; Dupont, 2012).

Connections listed above are not always optimally implemented in current OOFS. Rivers, atmosphere, and sediment exchanges are often introduced in a simplified way using climatologies or simplified exchanges. More refined interactions, including additional numerical modules or interannual observational data, are currently developing, and connections with surrounding systems should be considered for the construction of future systems.

9.2.4. Ensemble modelling

A forecasting system is literally designed to give an expectation of future conditions, having some knowledge of present conditions. The expectation is also a judiciously named statistic defined by the mean of all possible outcomes; for example, the expected primary production at a given location next week (time t_1) can be expressed as the mean of all possible values at the same time and location $<x(t_1)> = \int x(t_1)dx$. If we make next week's primary production a function of today's primary production $x_1=f(x_0)$, the function f() implicitly includes all the other variables than primary production at present time such as nutrients, solar activity, currents, etc. We obtain a new expression for the expected forecast value (using the notation <.> for the expected value) $<x(t_1)> = \int f(x(t_0))dx$. The function f() is unfortunately not a linear function because it represents the Michaelis-Menten equations (see Section 9.1.2.1), which after time integration become exponentials: if the concentration of plankton doubles today, you may expect a lot more than twice the plankton next week in a period of multiplicative growth. This means that one cannot swap the above integral and the f() function, even if $x_1 = f(x_0)$ is true, $<x_1>=f(<x_0>)$ is generally false and will ineluctably generate a biased expectation: too high or too low depending on the convexity of the f() function.

One general workaround for this problem is the use of an ensemble of simulations. Assuming that only a finite number of N possible outcomes is available, $<x(t_1)>$ becomes an arithmetic average instead of an integral: $<x(t_1)> \cong 1/N\ \Sigma(XN(T_1))$, with xn being a member of the ensemble: of the N possible outcomes, which are assumed independent from each other and identically distributed) If samples are like this, the arithmetic average will converge to the integral as N tends to infinity.

But why should one consider different possible outcomes when there is only one reality? The point is to manage uncertainties, which have more diverse origins in biogeochemical modelling than in physical or wave models, in particular the dependence on ocean physics is strong. Among the input data sources listed in Section 9.2.2, the following bear uncertainties that have an impact on biogeochemical model results:

- The seasonal restratification is critical. A too shallow mixed layer will confine the organisms near the surface and expose them to stronger lights than they should and exaggerate the bloom dynamics. A too shallow mixed layer will warm up too much and make the growth conditions artificially favourable. A late shoaling of the mixed layer in spring would lead to a delayed bloom in the simulation, leading to strong errors in surface Chla when comparing with observations.

- A good representation of winter mixing is also a desirable feature of the physical model, as it brings deep nutrients closer to the surface and makes them available for primary production.

- The ocean temperature influences the growth of microorganisms, so the simulated temperature should be accurate.

- The transport of nutrients from the rivers to the open ocean by ocean currents, or of any biological material from one oceanic region to another, requires accurate current simulations.

- The availability of light is fundamental for the ocean ecosystem. The amount of light reaching the surface of the ocean (i.e. how much light has been attenuated by the atmosphere, the clouds, the water, or sea ice) is uncertain.

- The initial conditions of the biogeochemical model are often based on very scarce climatologies of nutrients, some erroneous values may remain in the model during very long simulations.

- Nutrient inputs from rivers and atmospheric deposition are highly uncertain as well.

All the above are extrinsic source of errors, which can be accounted for by randomly perturbing various inputs of the biogeochemical model: perturbations of the downwelling shortwave radiations would account for uncertainties in light conditions, an ensemble of physical model outputs would account as well for the errors in the physical variables if the model is coupled offline. In the case of "online" coupling, the mixed layer depths can also be changed by adding perturbations to the surface winds and surface heat fluxes. There are various ways of generating random perturbations in 2 or 3 dimensions: a spectral method has been used in Natvik and Evensen (2003) and following works, but one could alternatively use an atmospheric ensemble prediction system or an empirical mode decomposition of atmospheric reanalysis data. The goal is to generate an ensemble of simulations, whose members differ slightly from each other because of the random perturbations they have received as input.

Intrinsic sources of errors have also been mentioned in Section 9.2.1. Among them, the BGC model parameters cannot be known with much certainty and can also be randomised. To do this, one needs to imagine their probability distribution, including their minimum and maximum admitted values. The random parameters may be fixed global values or values varying continuously in space (Simon et al., 2015) or discretely, according to designated provinces (the Longhurst provinces in Doron et al., 2011). Time-varying parameters also make sense since they may reflect neglected processes like population shifts. To this effect, an auto-regressive process is recommended in Garnier et al. (2016).

Other intrinsic sources of errors can be difficult to control, for example the noise caused by numerical advection schemes of the model or other model biases. If these are not too severe, it is desirable to emulate these uncontrollable errors by exaggerating the amplitude of other errors that can be controlled,

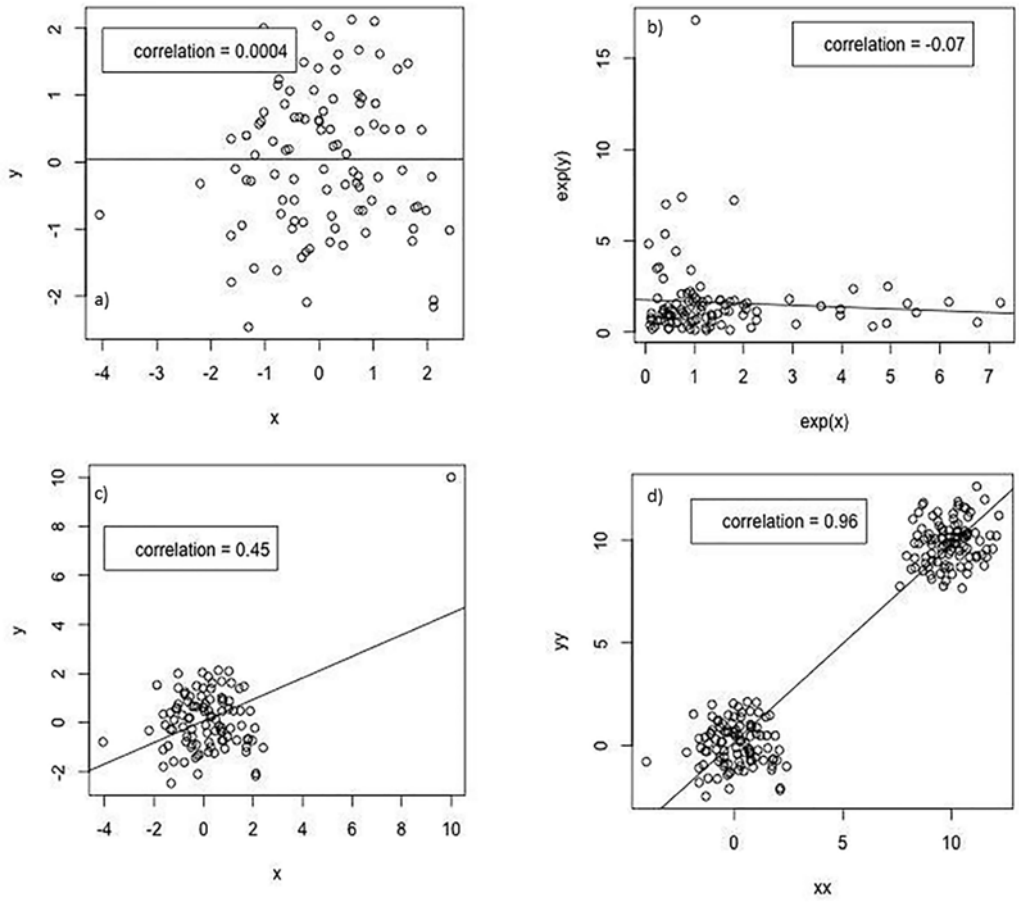


Figure 9.17. Scatter plots illustrating second-order statistics from various types of ensembles of size 100. a) two independent random Gaussian vectors x and y. b) their exponentials. c) same as (a) but adding one outlier at (10,10). d) mixture of independent Gaussian vectors, with an offset of 10. The correlations between the two variables are indicated in the legend.

e.g. increasing the level of noise in the wind forcing (extrinsic error) to compensate for a bias in the model mixing scheme (intrinsic). The preferred action, however, should be to correct the biases at their origin, if this is possible.

It is interesting to keep track of the perturbations applied, so that the differences between ensemble members can be explained by the sensitivity to the input parameters. A contrario, Garnier et al. (2016) also directly perturb the concentrations of biogeochemical tracers, in which case the differences between ensemble members can no longer be attributed to input parameters alone.

An ensemble of simulations is thus a way to obtain unbiased expectations, defined as a first-order statistical moment, but it also provides other higher order statistics as well. One statistic that is critical for data assimilation is the variance-covariance matrix, a second-order statistic. In particular, the statistics based on an ensemble can provide all empirical cross-covariances between observations and unobserved model variables, which are an essential ingredient of all data assimilation methods (Carrassi et al., 2018).

However, the variance and covariance estimated from ensembles are sensitive to outliers and may be wrongly estimated in case of ill-behaved ensembles. This is illustrated on Figure 9.17 with a synthetic example. Figure 9.17a shows the scatterplot of two independent Gaussian variables, x and y, that display a low correlation, as expected. The exponential of these values in Figure 9.17b shows a negative relationship due to the exponential stretching of randomly high values, which is not desirable neither for interpretation nor for assimilation. Figure 9.17c illustrates that the correlation can be very sensitive to the introduction of a single outlier. Figure 9.17d shows that a clustered ensemble can make the correlation artificially high, essentially making two hundred members equivalent to a two-members ensemble only.

9.2.5. Data assimilation systems

The assimilation of biogeochemical data into marine models aims at estimating the "true" value of biogeochemical quantities in real ocean ecosystems. These quantities are either key "states" of the ocean (e.g. the phytoplankton biomass) or "parameters" characterising the functioning of the ecosystem (e.g. the maximum phytoplankton growth rate). They are estimated by merging model guesses with field observations (e.g. model predictions and satellite observations of the phytoplankton biomass). Such merging weights the errors of both the model and the observation, looking for the "true" value that (ideally) lies in their proximity. Operational oceanography aims at estimating these "true" biogeochemical quantities to evaluate trends of ocean biogeochemistry in the past (in ocean biogeochemistry reanalysis), or to set initial values for biogeochemical model prediction in future forecasts.

The theory and methods behind data assimilation are described thoroughly in Chapter 4, while the biogeochemical model components have been described in the previous sections of this chapter. The following section provides a synthesis on how these ingredients can be combined in modern operational biogeochemical systems. Comprehensive reviews of the subject were published recently by Fennel et al. (2019) and Ford et al. (2018).

9.2.5.1. Biogeochemical state and parameter estimation

Most of the modern BGC OOFS apply DA to improve model simulations of biogeochemical state variables rather than biogeochemical parameters (Fennel et al, 2019). The main reason for this bias is the straighter link between model state variables and ecosystem indicators that interest end-users in the policy, management, and blue growth sectors. For example, the MFCs of the Copernicus Marine Service provide assimilative reanalysis and forecasts of nutrients, phytoplankton biomass and oxygen concentrations (linked to coastal productivity and eutrophication), and water acidity (pH, linked to ocean acidification and climate change). All these state variables are linked to the Sustainable Development Goal 14 (Life below water) and are targets of marine policy (e.g. the European Union Marine Strategy Framework Directive).

However, the variables targeted by BGC DA systems are not necessarily assimilated into the model. In fact, most of the above-mentioned centres assimilate ocean colour chlorophyll only, as a proxy of phytoplankton biomass, and none of them assimilates pH. It is assumed that a non-assimilated variable can be corrected towards its true value since it is linked to the assimilated variable, e.g. pH is improved through its relation to the phytoplankton biomass via photosynthesis/respiration that modify dissolved inorganic carbon (DIC) concentration and alkalinity in the water column, and thus pCO2 and pH. These corrections of non-assimilated variables can happen directly in the assimilative analysis step when using multivariate assimilation methods (Ciavatta et al., 2011). They can also happen indirectly during the model simulation of the ecosystem processes: in principle, an improved estimation of the phytoplankton biomass should quantify better the air-sea CO2 fluxes and hence their impact on pH. However, the improvement of non-assimilated variables is a strong assumption that needs to be thoroughly verified via comparison with independent datasets (see Section 9.2.6).

Some operational centres use BGC DA to estimate biogeochemical model parameters, on their own or concurrently with the model state variables (e.g. in a multivariate analysis configuration). For example, the Arctic MFC estimates rates of phytoplankton growth and mortality, and this improves the simulation of the phytoplankton biomass that is a target variable of the operational system (Simon et al., 2015). The parameters can be estimated as variables in time and space, to somehow represent the variability of the real system

which cannot be formulated in the mechanistic equations of the model. For example, the variability of the phytoplankton species that are represented in biogeochemical models are often forced into few functional groups. In practice, the spatial-temporal variability of a given biogeochemical parameter is often represented as a random variable, with predefined statistical distribution. Its fluctuations are computed through the minimization of a cost-function between model prediction and field observations of a state variable, which is linked to the parameter and assimilated into the model. BGC DA for parameter estimation has an enormous potential to improve our understanding of marine ecosystems, their model representation, and the operational prediction of target variables. However, it is also challenging, mainly due to the scarcity of data to define realistic statistical distributions for the parameter variability and assess the reliability of the estimated parameter fluctuations. Schartau et al. (2017) provided an excellent review of these opportunities and challenges.

9.2.5.2. Assimilated observational products

Most of the modern BGC OOFS assimilate ocean colour Chla into their model systems (Fennel et al., 2019). That is because this satellite product: i) quantifies the biomass of a central component of biogeochemical models (phytoplankton); ii) provides data that are generally synoptic (~100 km), high resolution (~100 m), and frequent (~daily); and iii) has a timely and free access (e.g. through the Copernicus Marine Service; [20]). A thorough discussion on the use of ocean colour in biogeochemical modelling and assimilation is provided in the report of the IOCCG (IOCCG, 2020). Here it is worth mentioning that, after the seminal assimilation of ocean colour by Ishizaka (1990), biogeochemical reanalyses were produced by assimilating ocean-colour total Chla in the global ocean (Nerger and Gregg, 2008), in an ocean basin (Fontana et al., 2013), and in coastal and shelf-seas ecosystems (Ciavatta et al., 2016). More recent contributions include the decadal global ocean ecosystem reanalyses by Ford and Barciela (2017), obtained by assimilating different ocean colour products for 1997 to 2012, and the one by Gregg and Rousseaux (2019), who estimated global trends of primary production by assimilating ocean colour for 1998-2015. Besides the well-established assimilation of total Chla from ocean-colour (e.g. Hu et al., 2012), innovative applications have assimilated surface ocean colour products for: spectral diffuse attenuation coefficients (Ciavatta et al., 2014), size-fractionated Chla and POC (Xiao and Friedrichs, 2014), remote sensing reflectance (Jones et al., 2016) and both phytoplankton functional type Chla and spectral absorption (Ciavatta et al., 2018 and 2019; Skakala et al., 2018 and 2020; Pradhan et al., 2020). Surface data of partial pressure of CO2 (pCO2) from ships of opportunity were used in the reanalysis of air-sea CO2 fluxes in the global ocean (While et al., 2012).

20. https://marine.copernicus.eu/

Biogeochemical data are sparse for the ocean interior, but they can be useful to constrain vertical gradients that are extremely important in the functioning of marine ecosystems. For example, biogeochemical simulations were improved by assimilating vertical observations of nutrients, oxygen, and pCO2 data at fixed stations (Allen et al., 2003; Torres et al., 2006; Gharamti et al., 2017). The increasing number of autonomous underwater vehicles and floats observing biogeochemistry in the global ocean is an opportunity for the development of operational oceanography (see also Section 9.2.2). The assimilation of such data in the water column can complement the assimilation of ocean colour at the surface of the ocean. For example, glider data of Chla and POC were assimilated by Kaufman (2017), while Skakala et al. (2021a) assimilated glider Chla and oxygen data along with ocean colour data in an operational model of the European North West Shelf Seas. Recently, the assimilation of BGC-Argo float data led to improvements in the simulation of subsurface biogeochemistry in regional seas (Verdy and Mazloff, 2017; Wang et al., 2020), as well as in the global ocean (Carroll et al., 2020). OSSE analyses have shown the potential of improving the ocean biogeochemical simulations by combining the assimilation of the planned 1000 BGC-Argo fleet with ocean colour assimilation, with both variational data assimilation methods (Ford, 2021) and stochastic ensemble approaches (Germineaud et al., 2019). The Mediterranean MFC pioneered the assimilation of the BGC-Argo float for operational oceanography (Cossarini et al., 2019). This application is demonstrating remarkable advantages for the prediction of the subsurface phytoplankton dynamics and biogeochemistry, with respect to the assimilation of ocean colour alone. It also pointed out the current main challenges in using the BGC-Argo float data operationally: i) the availability of quality-controlled data in near-real time; ii) the relatively low number of floats available currently, which implies that the impact of their assimilation is spatially constrained; and iii) potential biases between the assimilated float and satellite data (e.g. the Chla concentrations derived for remote sensitive reflectance and in-situ fluorescence).

9.2.5.3. Biogeochemical data assimilation methods

The general theory and application of data assimilation methods were presented in Chapter 5. For the assimilation of biogeochemical data, current operational systems are using two methods (Fennel et al., 2019; Moore et al., 2019):

a. Ensemble methods, which use an ensemble of ocean model simulations or states to represent the evolution of the physical and biogeochemical state variables and their uncertainty.

b. Variational methods, which correct the model simulation towards the observation by minimising the differences between the observation and the model estimate of the variable.

Hybrid ensemble/variational assimilation methods have been applied successfully with physical ocean models (e.g., Storto et al., 2018) and are currently being developed for the assimilation of biogeochemical data in operational systems of the Copernicus Marine Service (EU H2020 SEAMLESS project: [21]).

There is no "best" method for the assimilation of biogeochemical data. The choice depends mainly on: i) the target variable (or parameter) of the assimilative simulation; ii) the data being assimilated; and iii) the computational resources, which can become a major issue when using biogeochemical models with a large number of variables. For example, an ensemble method might be preferable if the target variable (e.g. nitrate) is different from the assimilated variable (e.g. ocean colour Chla) because one can exploit multivariate analyses that take the dynamical model error covariances into account. If the number of CPUs is a concern, efficient variational methods might be the best choice, if adequate information about the model error covariances is available.

As far as ensemble methods are concerned, since the introduction of the original EnKF (Evensen, 1994), different flavours of the filter have been developed (Vetra-Carvalho et al., 2018) and applied with operational biogeochemical systems (Fennel et al., 2019). For example, both the reanalysis system of the Arctic Ocean (Simon et al., 2015) and the operational system of the Great Barrier Reef (Jones et al., 2016) use the DEnKF (Sakov and Oke, 2008). In the Baltic MFC, work is in progress to apply the local ESTKF (Nerger et al., 2012), while the Global MFC is based on the SEEK (Pham et al., 1998). However, the propagation of an ensemble of model states implies a high computational cost. To ensure that the EnKFs perform adequately with affordable ensemble sizes (i.e. between 10 and 200), practical adaptations like "localization" have been adopted (Houtekamer and Mitchell, 1998). Localization approaches correct the model simulation towards the observation just around the point where the observation was taken. "How much around" (i.e. the localization scale) depends also on the spatial variability of the variable that is observed.

Examples of variational methods for biogeochemistry used by some operational centres include: the 3D-Variational assimilative system for the European North West Shelf Seas (Skakala et al., 2018) using NEMOVar (Mogensen et al., 2009; Waters et al., 2015) and for the Mediterranean Sea using 3DVarBio (Teruzzi et al., 2014 and 2019); the 4D-Variational system of the CCS (Song et al., 2016). In all the above cases, the assimilated variable is ocean colour Chla concentrations, but a limited number of other model variables are also updated by means of functional links such as background Chla-to-nutrients ratios of the phytoplankton cells.

A particular issue for biogeochemical data assimilation methods is the non-Gaussianity of the distributions of the biogeochemical variables (Campbell, 1995), which is related to the non-linearity of the ecosystem processes. In fact, most of the traditional methods assume that these distributions are Gaussian. The use of logarithm of the concentrations, in particular for Chla assimilation (Nerger and Gregg, 2007) and Gaussian anamorphosis (Bertino et al., 2003), has been demonstrated to handle the issue by bringing distributions closer to Gaussian before the assimilation of the data. This approach is currently exploited in operational systems of the Copernicus Marine Service, e.g. in the centres for the European North West Shelf Seas, Arctic and Global oceans (Simon et al., 2015; Skakala et al., 2018; Lamouroux et al., in prep.).

9.2.5.4. Current challenges and opportunities

State-of-the-art operational centres are using BGC DA to provide better model output products to their users. It is expected that this use will expand further in the future thanks to current research and developments that are addressing the BGC DA challenges and opportunities described below (see also Fennel et al., 2019, and the EU H2020 SEAMLESS project ([22]) specifically dedicated to the advancement of operational biogeochemical data assimilation systems).

Before applying any BGC DA method, the physical-biogeochemical models at hands need to be improved as much as possible, e.g. through implementation of the most relevant processes, improved parameterizations, corroboration of equations, and simulation by using laboratory and field data. In fact, biogeochemical data assimilation cannot fix (and actually might deteriorate) any systematic flaw of the applied ecosystem models (Ciavatta et al., 2011).

It is expected that the integrated assimilation of data from the expanding fleets of in-situ autonomous observing systems (e.g. BGC-Argo floats in the open ocean and gliders in shelf-seas and coastal areas), along with the traditional surface ocean colour data, will make possible to constrain better a larger number of model variables and parameters of operational models (Cossarini et al., 2019; Skakala et al., 2021a).

In current applications, the assimilation of physical data into ecosystem models can cause the deterioration of the biogeochemical simulations due to the breaking of physical balances and of their consistency with the biogeochemical fields (Anderson et al., 2000). In models of the equatorial ocean, the assimilation of temperature and salinity profiles, or sea surface height, can perturb the balance between pressure gradients and the wind stress, generating unobserved currents and spurious vertical velocities (Waters et al., 2017; Park et al., 2018). In turn, this can result in unrealistic upwell-

21. https://seamlessproject.org/

22. www.seamlessproject.org

ing of nutrients and excessive recreation of the water column, deteriorating biogeochemical model products (e.g. oxygen and primary production). The combined assimilation of physical and biogeochemical data is a promising approach to address the above issue, and preserve the consistency between the physical and biogeochemical simulations (Anderson et al., 2000; Ourmières et al., 2009; Song et al., 2016; Yu et al., 2018). Using bio-optical modules, which provide feedback from biology to ocean physics in "two-way" coupling interactions, models are expected to preserve even better such consistency, in both simulation and assimilation steps of operational systems (Skakala et al., 2021b). The opportunity for the combined assimilation of physical and biogeochemical data is increasing along with the growing number of BGC-Argo floats and gliders mounting multivariate sensors in the ocean (Skakala et al., 2020).

The steady expansion of computing capability will facilitate the use of ensemble methods (including hybrid ensemble-variational methods) to better represent the dynamics of the biogeochemical model errors. Nevertheless, this evolution should be accompanied by the use of new stochastic ensemble generation methods that can represent the actual model uncertainty (Santana-Falcon et al., 2020), and the careful consideration of potential non-linearity/non-Gaussianity issues that can weaken the applicability of traditional data assimilation methods. To address these issues, new DA methods such as particle filters (van Leeuwen, 2010) have been applied to coupled physical-biogeochemical models (Mattern et al., 2013) and might be used in operational systems in the future.

Finally, Artificial Intelligence/Machine Learning methods have supported data assimilation with geophysical models and will likely become relevant components of future operational biogeochemical data assimilation systems (Mattern et al., 2012).

9.2.6. Validation strategies

The validation methodology is built upon four classes of metrics that have been defined by the GODAE/OceanPredict community (Figure 4.30) to monitor the quality of ocean analyses and forecasts in physics (Section 5.7) and are used and supported by the broader biogeochemical community. These metrics gather a complete range of statistics and comparisons in space and time to properly assess the consistency, representativeness, accuracy, performance, and robustness of ocean model outputs. They are classified as follows (for a more detailed description see Hernandez et al., 2009 and Chapter 4):

- **Class 1:** metrics aim to provide a general overview, they are typically spatial maps or vertical profiles.
- **Class 2:** metrics correspond to virtual moorings or sections of the model domain.
- **Class 3:** metrics are derived quantities, such as integrated quantities.
- **Class 4:** metrics are model-observation match-ups products.

Based on this methodology, the validation strategy of biogeochemical operational systems consists of two phases: i) the near-real time evaluation of the forecast accuracy; and ii) the delay mode evaluation of the model system.

9.2.6.1. Near-real time evaluation

The NRT validation aims to provide information about the quality of the forecasts and relies on the availability of NRT observations (e.g. data from satellite and from autonomous underwater sensors such as BGC-Argo floats, BGC-gliders, and moorings equipped with biogeochemical sensors). A validation is defined as semi-independent (independent) when the observations are (not) assimilated in a sequence of analysis and forecast cycles. In fact, an observation from a continuously recording sensor, even if not yet assimilated, shares some level of correlation with already assimilated observations from the same sensor.

The forecast validation is commonly based on temporal and spatial match-ups of forecast model outputs and observations (i.e. GODAE Class 4 metrics), and on the computation of statistical skill metrics such as average difference (also referred to as average misfit or bias), average absolute difference, RMS Difference (RMSD), correlation index, and model efficiency (Stow et al., 2009). Skillfulness of forecasts can be compared in terms of persistence (i.e. comparison with previous day forecast) or with skill performance against a reference climatology. Skill statistics are often reported for various forecast lengths (i.e. number of days in the future).

Two examples are presented in the following figures. Figure 9.18 shows model analysis, six days of forecast and compare surface Chla model estimates to satellite observations for the European North West Shelf Seas system. Successive daily forecast values quickly diverge from the satellite product during spring and summer months, highlighting the strong effect of data assimilation during the production period. During winter, the satellite coverage decreases and the ocean colour error increases, inducing a negative forecast bias.

Figure 9.19 shows statistics for 1st and 6th forecast day in the Arctic Ocean. The onset of the spring bloom in the model is significantly delayed, but from the middle of May onwards, the model results are close to the observations. The quality barely changes as the length of the forecast period increases, except during the spring bloom (the first weeks of the time series) in which the bias is significantly smaller for a forecast range of one day, suggesting that, at this stage, the model is unable to support increased concentrations after the assimilation events.

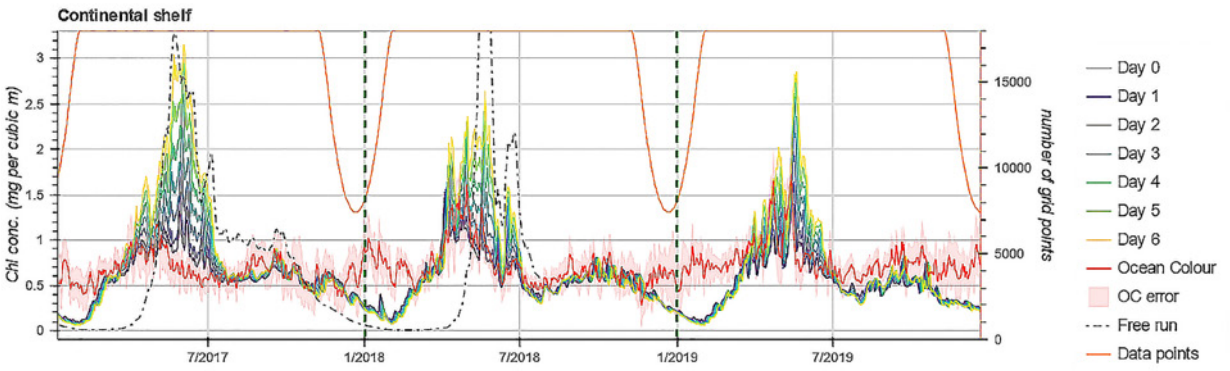


Figure 9.18. Time series of surface Chla concentration for European North West Shelf Seas average. Day 0 is the analysis day, with assimilation of satellite Chla, and days 1-6 are forecast days. Satellite ocean colour values are shown in red for comparison and error in the pink shaded area. The number of grid point matchups is shown in orange (from McEwan et al., 2021).

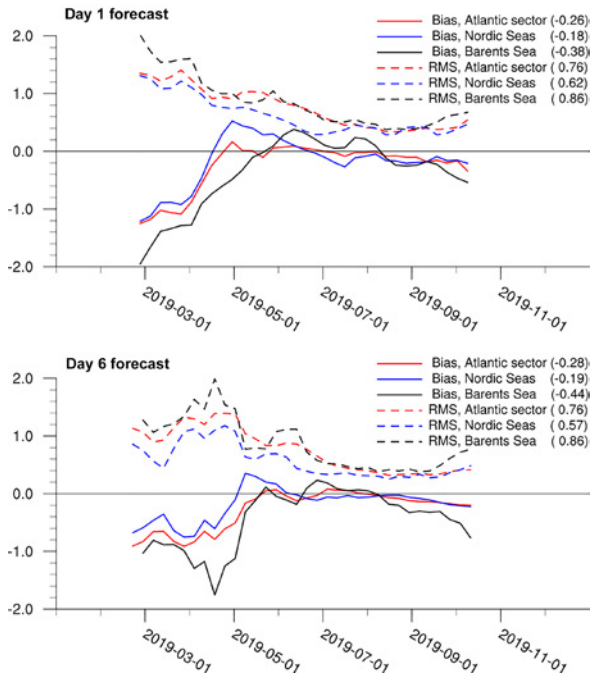


Figure 9.19. Time series for bias and root mean square (RMS) differences between the Arctic Ocean model system and ocean colour satellite for 1st (top) and 6th (bottom) forecast day. Statistics are given for the various regions, log10 transformation has been applied (from Melsom and Yumruktepe, 2021).

A different class of metrics can be used to evaluate the capacity of the forecast system to reproduce specific events, such as algal blooms. In this case, the skill metrics are based on a binary discriminator test with a threshold (i.e. greater or lower than a given value of Chla concentration) and a yes/no decision. For example, the ROC (Brown and Davis, 2006) compares two independent information sets (i.e. forecast and observation) with respect to a threshold value. For each couple of yes/no decisions there are four possible outcomes, either correctly positive or correctly negative, and two model failures for incorrectly positive and incorrectly negative. Results of such metrics are plotted in contingent tables (Stow et al., 2009).

An example of the use of the ROC to characterise Chla in terms of events is presented in Figure 9.20, using the Mediterranean Sea system. The threshold is defined as the 75th percentile of surface concentration and identifies surface bloom occurrence.

Since biogeochemical variables are often not Gaussian distributed (e.g. surface Chla distribution resembles a log-normal density distribution), forecast skill performance metrics can be computed on log-transformed data or using non-parametric statistics, for example median of the misfit (i.e. model minus observation) instead of bias, interquartile range of the misfits instead of RMSD, and Spearman correlation instead of Pearson one. However, while data transformation (such as the log-transformation) preserves the statistical robustness of metrics, it results in metric values that may be difficult to understand by users, thus reducing the benefit of the validation information (Hernandez et al., 2009).

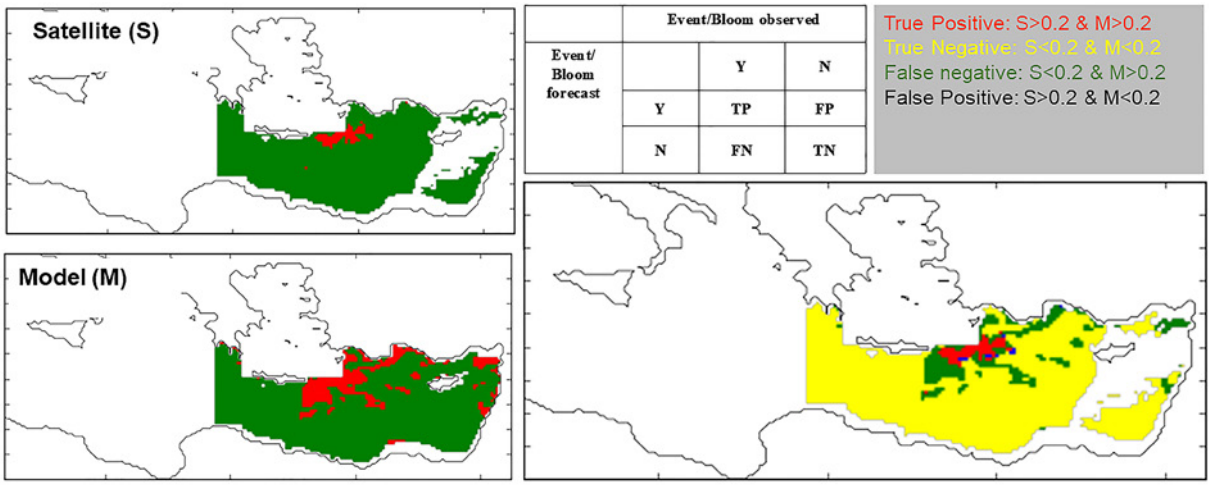


Figure 9.20. Time series of surface Chla concentration for European North West Shelf Seas average. Day 0 is the analysis day, with assimilation of satellite Chla, and days 1-6 are forecast days. Satellite ocean colour values are shown in red for comparison and error in the pink shaded area. The number of grid point matchups is shown in orange (from McEwan et al., 2021).

Real time skill statistics are reported in web pages which are continuously updated (e.g. the validation dashboard of the Copernicus Marine Service: 23). Indeed, time series of the validation metrics monitor the quality of the operational biogeochemical system and alert for quality degradation of the model outputs. Possible deviation from the nominal quality of the forecast products, which is specified in the delay mode validation, might be due to model failure to capture specific events, degradation of upstream input data (e.g. assimilated observations), model internal biases, but also to the day-to-day fluctuation in the number of available observations.

9.2.6.2. Delay mode evaluation

The DM validation conveys a more comprehensive and detailed evaluation of the model capability to reproduce the spatial and temporal scales of variability of marine biogeochemistry. DM validation assesses the reliability of the model results considering the user needs and requirements, measures the strengths and weaknesses of the model system for future developments, and defines the nominal quality level to which the forecast skill performance can be compared (Hernandez et al., 2018).

9.2.6.2.1. Common graphical representations

Results of the model performances assessment are generally provided in a variety of graphical representations that can be complementary each other, most common representations are:

- Spatial maps represent the spatial distribution of a given variable and highlight the model's ability to reproduce global patterns, spatial gradients, and basin inter-difference. The bias and RMSD maps between predicted and observed values identify the regions of high and low model uncertainty.

- Scatter plots compare the predicted values with the observed values in the form of a collection of pair-values (i.e. points in a model vs observation graph). If the points are coloured, one additional information can be displayed. Scatterplots are useful to identify relationships between the predicted and observed values.

- Vertical profiles compare the vertical structure of the predicted values with the observed values: surface values, vertical gradient, and deep content. The shape of the profiles gives indications of the physical and biogeochemical dynamics at work.

- Time series graphs represent the evolution of predicted values with the observed values as a function of time. Such representation allows analysis if temporal dynamics (such as seasonal variability, interannual variability or trends) are captured by the model.

- Hovmöller diagrams are latitude/longitude/depth versus time diagrams displaying the evolution of a variable. They are more powerful than the time series graphs because they offer an additional dimension, allowing to study how models reproduce spatial or vertical dynamics over time.

23. https://pqd.mercator-ocean.fr/

- Taylor diagrams display simultaneously information on model-observations skill for three metrics (Taylor, 2001): 1) the Pearson correlation coefficient, 2) the RMSD, and 3) the SD. RMSD and SD are usually normalised (RMSD and the model SD are divided by the SD of the observations) to represent all metrics with different units into a single diagram (normalised Taylor diagram). The Pearson correlation coefficient between the model and the observations is given by the azimuthal position. The normalised SD is proportional to the radial distance from the origin. The normalised RMSD is proportional to the distance from the reference point. The observational reference is indicated along the x-axis and corresponds to the normalised SD and correlation equal to 1. Such diagrams are used to compare different model versions or to summarise the model performance for different variables in a single and compact diagram (Jolliff et al., 2009).

9.2.6.2.2. Evaluation of different system versions

In the frame of the continuous improvement principle, any upgraded and novel version of an operational biogeochemical system should show advancements with respect to the previous one in terms of model characteristics (e.g. addition of new modelled variables and processes) and quality of the results. Updates of model formulations and upstream input data contribute to reduce the system uncertainty with respect to a standard skill performance framework allowing versioning comparison. Figures 9.21 and 9.22 show how metrics can be used to compare different versions of a system.

Figure 9.21 compares daily surface Chla for two model versions of the European North West Shelf Seas system using regionally-averaged time series (GODAE Class 4 metrics). The new product (V11 in Figure 9.21) is constrained by data assimilation while the previous product (V10 in Figure 9.21) was not. The new version shows a better match with satellites during the growing season, with lower summer peak and earlier spring bloom, although there are differences among regions.

In Figure 9.22, the Taylor diagram summarises the quality improvement for different system versions of the Irish-Biscay-Iberia MFC. Chla, nutrients, oxygen, and carbon variables are compared to ocean colour, WOA and GLODAP (GODAE Class 4 metrics). The evolution of the system shows an improvement in almost all variables, and particularly in carbon-related variables. This improvement is due to more realistic initial and boundary conditions in the latest version of the system.

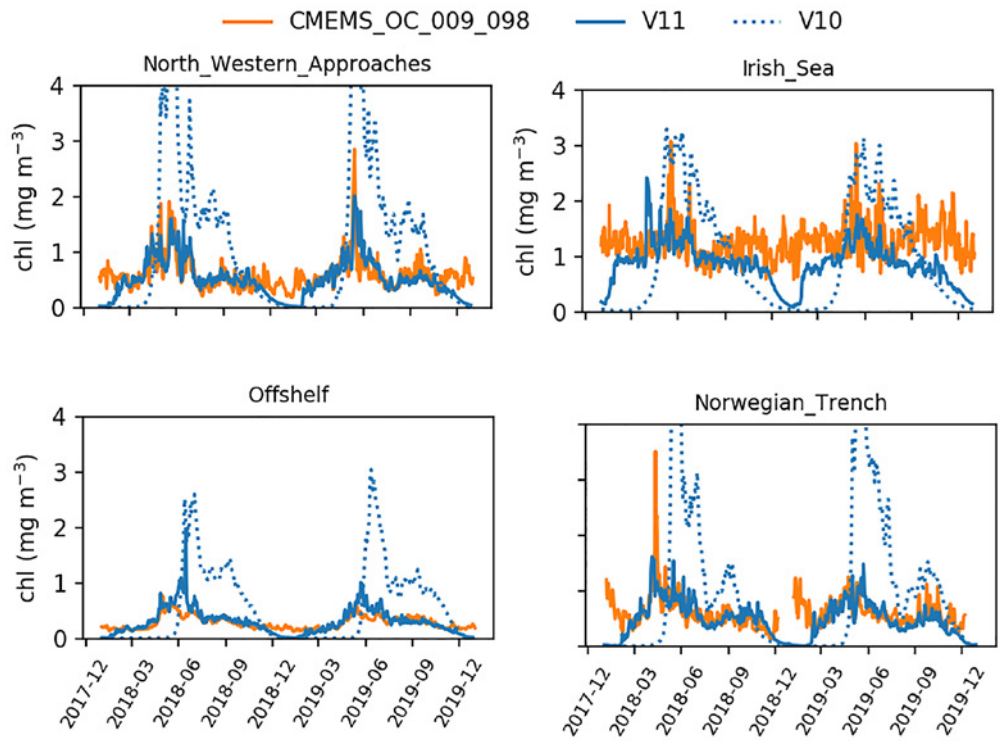


Figure 9.21. Time series of daily surface Chla for regions of the European North West Shelf Seas, from the new product (V11), the previous version (V10), and ocean colour satellite (from McEwan et al., 2021).

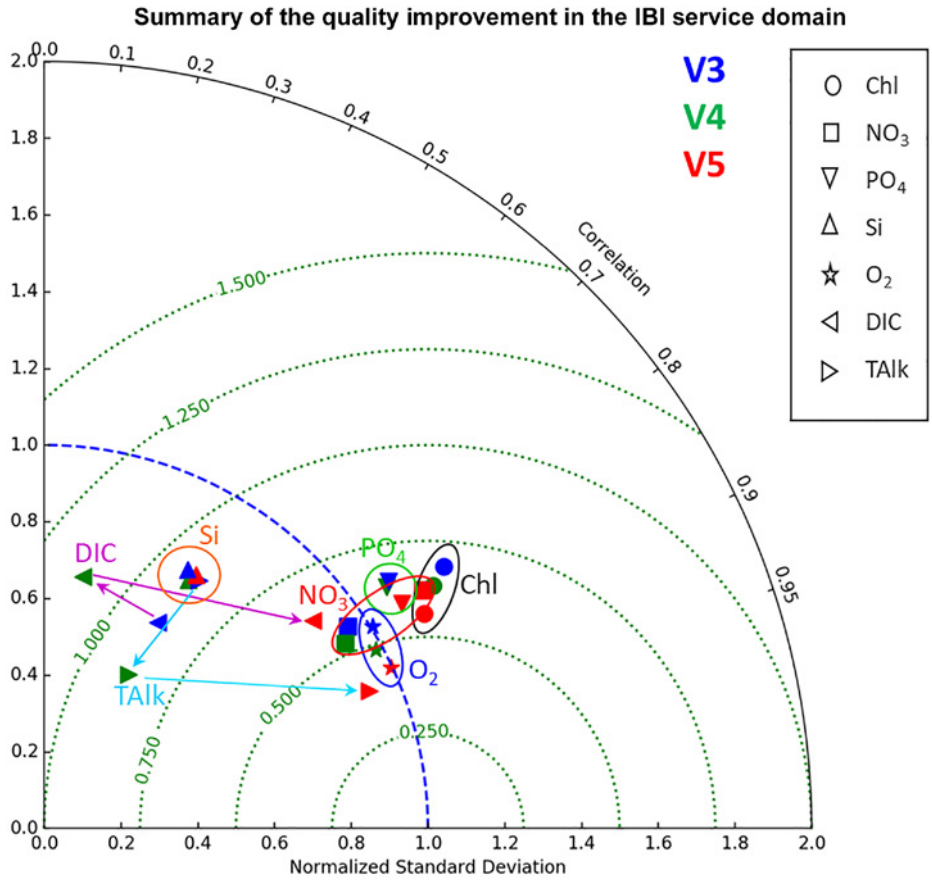


Figure 9.22. Taylor diagram summarising the quality improvement of the operational system of the Irish-Biscay-Iberia MFC (part of the Copernicus Marine Service).

9.2.6.2.3. Spatial and temporal evaluation

The DM validation is commonly built to test the pre-operational system for a medium/long simulation using higher quality observation datasets. They can include the same observation data of the NRT validation but characterised by a higher quality check (e.g. reprocessed ocean colour product) and an additional number of historical in-situ data collections (e.g, World Ocean Database, SOCAT, EMODnet data collection) that, because of the delay mode quality check, become available a certain time after their acquisition time.

Chla derived from remote sensing is a major dataset for BGC OOFS. It is extensively used in DM validation to validate the spatial and temporal structures. Figure 9.23 shows the annual average distribution of Chla from the model and satellite observations (i.e. GODAE Class 1 metrics). The large-scale structures present a good agreement, i.e. the main biogeographic provinces of Longhurst (1998) including oligotrophic gyres (low levels of chlorophyll in the centre of the basins), Antarctic Circumpolar Current, tropical band, Eastern Boundary Upwellings, are well reproduced. Differences at the regional spatial scale are found along the equatorial band, in the southern high latitudes, and in coastal regions as highlighted by the scatterplot (Figure 9.24). The distribution of points shows good estimations in the open sea (for depths higher than 1000 m) and underestimations in shallow waters (when bathymetry is lower than 1000 m).

Seasonal cycle and interannual variability can be analysed using Hovmöller diagrams. Figure 9.25 shows the seasonal cycle of Chla in the North Atlantic, from the Global Ocean system of the Copernicus Marine Service. The main features reproduced are: i) a bloom in spring when the mixed layer, rich in nutrients, shoals (light limitation); ii) a decrease of Chla concentration in summer due to a thin mixed layer very poor in nutrients (nutrient limitation); iii) a secondary bloom in autumn when the mixed layer is deepening and nutrients are transported in the euphotic layer; iv) a period of weak production (light limitation) in winter; and v) a marked seasonal cycle in the extension of the subtropical gyre (retraction in winter and extension in summer). The interannual variabili-

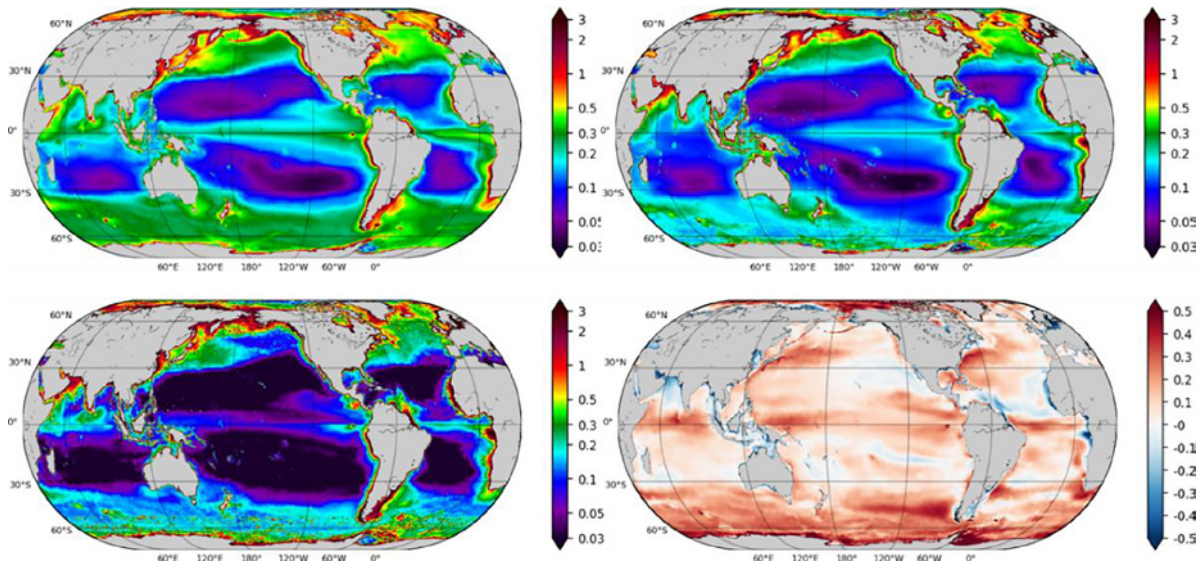


Figure 9.23. Annual mean of surface Chla averaged over the period 2009-2018 (mg Chl m^{-3}). Top left: model. Top right: satellite L4 observations. Bottom left: RMSD between model and satellite observations. Bottom right: log bias (i.e. mean difference of log) between model and observations (from Lamouroux et al., 2019).

ty of the south boundary of the oligotrophic gyre (i.e. the area between 30°N to 40°N) is also well reproduced by the model.

Long-term oceanographic monitoring stations are invaluable platforms to investigate temporal and spatial scales of BGC variability and assess BGC and ecosystem models. An example is the BATS in the Sargasso Sea, situated in the North Atlantic subtropical gyre. Figure 9.26 compares the Chla modelled and measured at this station. The model predicts reasonably well the subsurface Chla maximum, with concentrations slightly higher than in BATS data. The model catches the seasonal cycle, with a bloom during the deepening of the mixed layer in winter. In summer, the production in the mixed layer is more limited and is mainly due to the local remineralization of organic matter (regenerated production).

Observations for a large number of variables are also available in historical in-situ collections (e.g. nutrients like nitrate, phosphate, ammonium, silicate, iron; and carbonate system variables like dissolved inorganic carbon, alkalinity, pH, pCO2, biomass for phytoplankton and optical quantities) contributing to enrich the state validation framework embracing multiple features of the biogeochemical model.

Figure 9.27 presents a multivariate GODAE Class 1 quantitative comparison between model average vertical profiles and the reference EMODnet climatological profiles in the North West Mediterranean sub-basin. The model reproduces the average values and shape of the profiles; modelled profiles are within the range of variability of the climatological profiles.

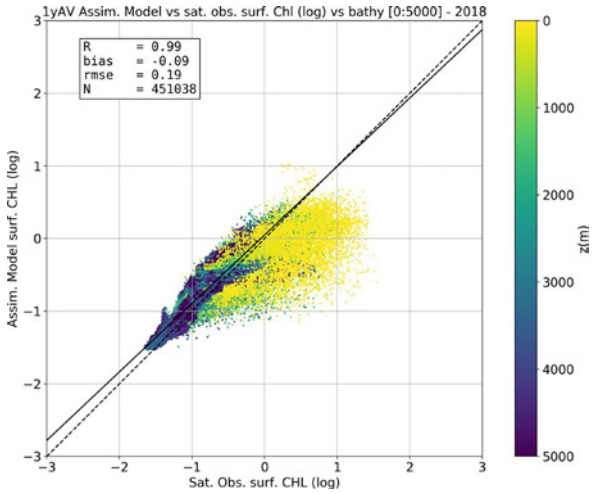


Figure 9.24. Scatterplot comparison of 2018 annual averaged surface Chla concentration for the model vs satellite observations. The colorbar represents the bathymetry (m), from shallow (yellow) to deep water columns (dark blue). The dashed line is the line 1:1, the plain line is the least square regression fit within the data. The correlation coefficient R, the bias, the RMSD (referred to as rmse) and the number of points N are computed on the log10-transformed space (from Lamouroux et al., 2019).

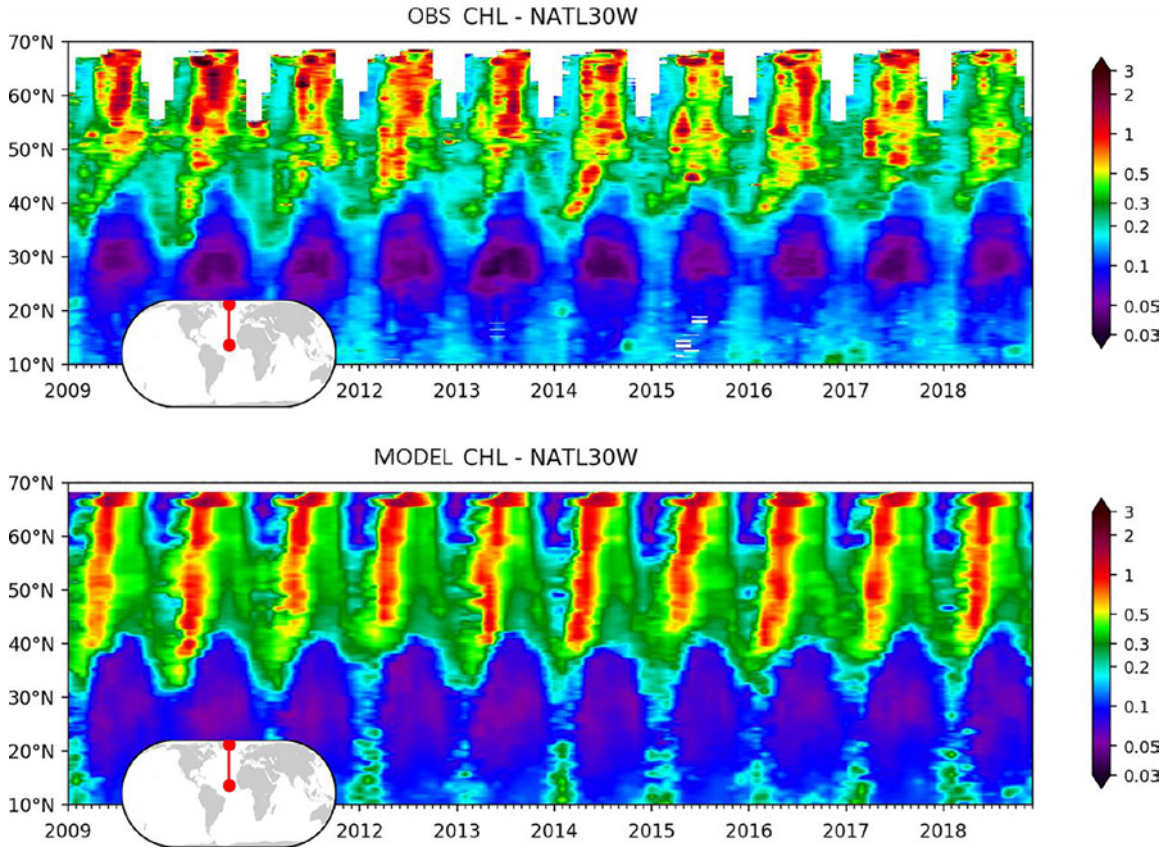


Figure 9.25. Hovmöller diagram (latitude versus time) of surface Chla concentration on 2009-2018 period computed with monthly mean fields. Top: model. Bottom: satellite observations (from Lamouroux et al., 2019).

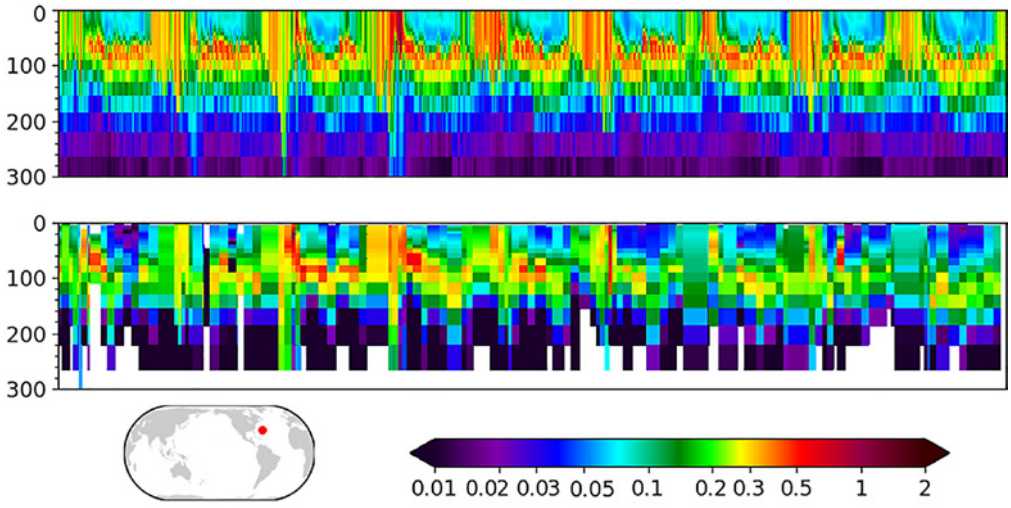


Figure 9.26. Hovmöller diagram (depth versus time) of Chla concentration (mg Chl .m-3) in the layer 0-300 m at BATS station, over the period 2008-2017. Top: model. Bottom: bottle data at BATS station (from Lamouroux et al., 2019).

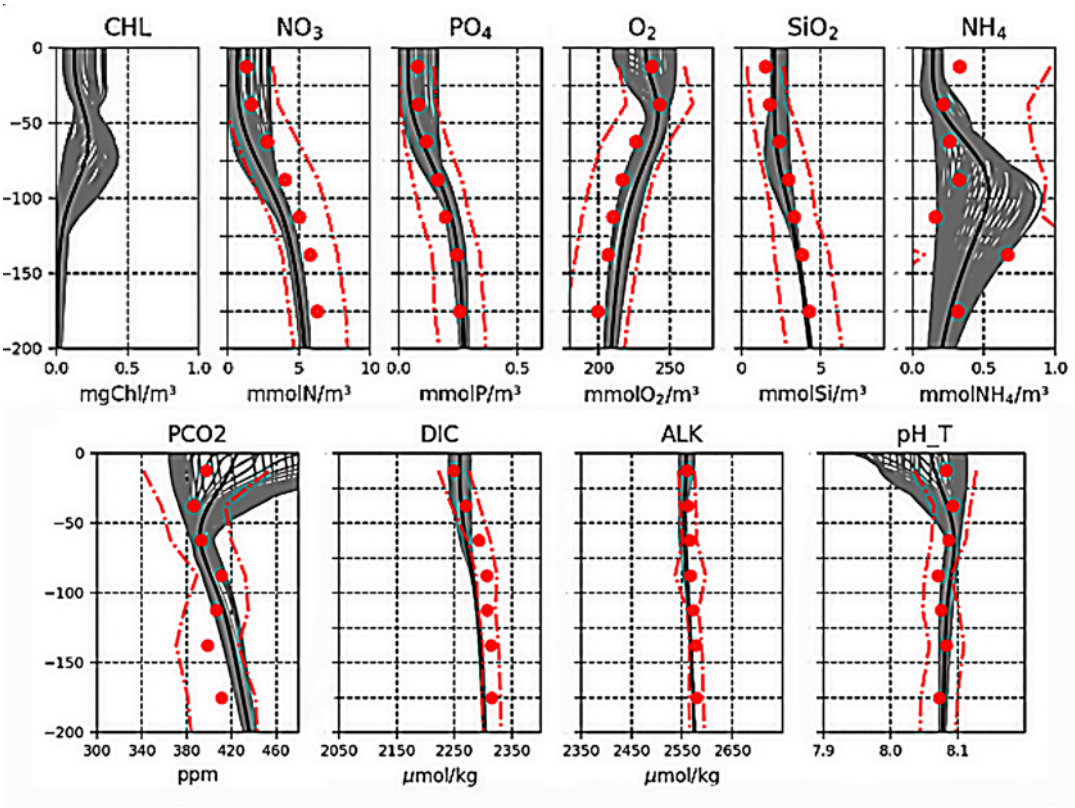


Figure 9.27. Comparison between weekly (grey lines) and annual (black lines) vertical profiles from the model run for North West Mediterranean sub-basin in 2019 (part of the Copernicus Marine Service) and climatological profiles of nutrients, dissolved oxygen, and carbon variables retrieved or derived from EmodNET dataset (red dots for means and dashed lines for standard deviations) (from Salon et al., 2019; Feudale et al., 2021).

9.2.6.2.4. Process-oriented evaluation

Besides the already mentioned direct skill error calculation (e.g. bias, RMSD) and pattern assessments (e.g. spatial correlation between model and observational maps), DM validation is enriched by process-oriented metrics (i.e. quantities derived from state variables that describe the results of particular processes) and theoretical derived quantities, such as stoichiometric indicators N:P, DOC:POC, Chla:POC, which contribute to assess the fit-for-purpose of the model functioning. Among process-oriented metrics, it is worth mentioning those deriving from the use of the continuously growing amount of available BGC-Argo floats and glider profiles. Metrics are based on the depth, slope, and amplitude of several particular biogeochemical features, such as the deep Chla maximum, nitracline, and oxygen minimum zones. They are associated with the biological carbon pump, the air-sea CO2 flux, oceanic pH, oxygen levels, and provide an innovative framework that evaluates the model capability to reproduce the interaction of physical (e.g. vertical mixing) and biogeochemical (e.g. phytoplankton growth and uptake) processes that shape variable vertical profiles (Salon et al., 2019; Mignot et al., 2021).

These metrics are currently used for DM validation but could also be easily implemented for NRT validation by routinely comparing the forecast skill with pre-operationally defined seasonal benchmarks and thus highlighting possible anomalies. For example, Salon et al. (2019) used such metrics to evaluate the system of the Mediterranean Sea (Figure 9.28), while Mignot et al. (2021) applied them to evaluate the system of the Global Ocean (Figure 9.29 and 9.30), both part of the Copernicus Marine Service.

Figure 9.28 shows how the time evolution of the vertical profiles matches up with the observations as well as several quantitative metrics along the corresponding float trajectory in the Mediterranean Sea. Temporal succession of the winter vertically mixed blooms, the onset, the time evolution, and the depth of the DCM, which typically establishes during the

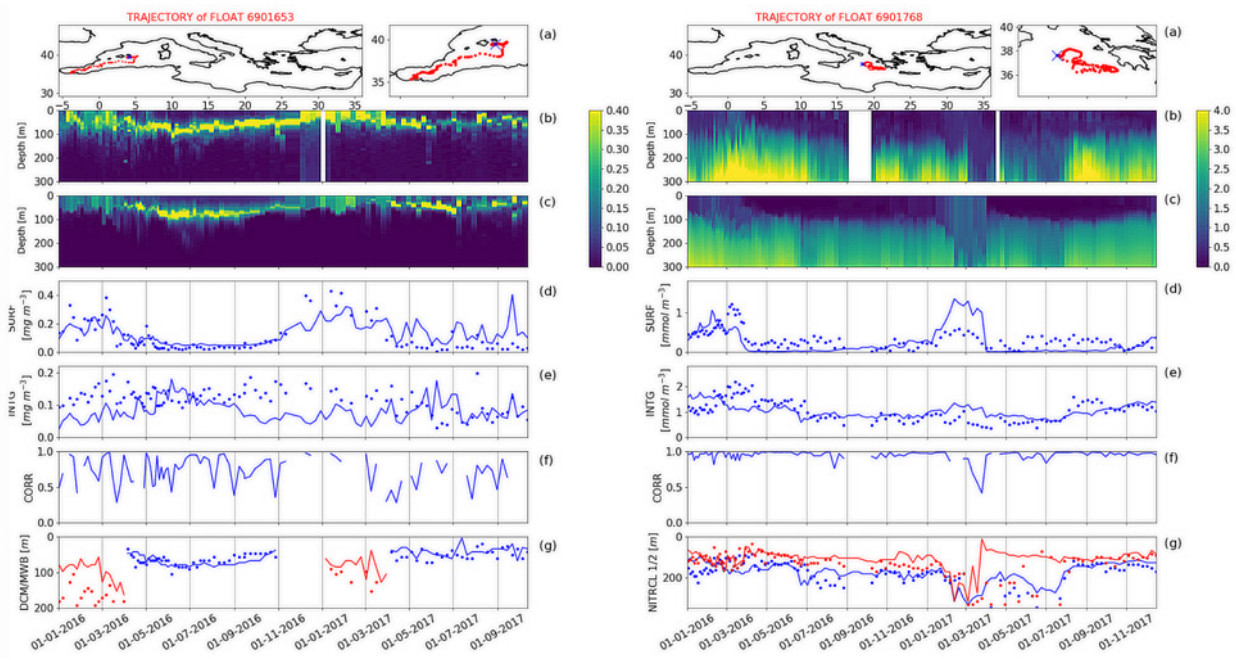


Figure 9.28. Time evolution of two BGC-Argo floats in Mediterranean Sea for Chla (left) and nitrate (right). a): trajectory of the BGC-Argo float; b): Hovmöller diagrams (depth versus time) of Chla and nitrate concentration from float data; c): model outputs matched-up with float position. Metrics for model (solid line) and float data (dots): d): surface concentration; e): 0–200 m vertically averaged concentration; f): correlation between vertical profiles; g): depth of the deep chlorophyll maximum (blue) and depth of the mixed layer bloom in winter (red) to the left, and depth of the nitracline (2 calculation methods) to the right (from Salon et al., 2019).

stratified season, are consistent in the Western Mediterranean Sea (Figure 9.28, left). The analysis is completed by Chla profiles, nitrate content, and nitrate-based metrics (Figure 9.28, right) that allow to evaluate the key coupled physical–biogeochemical processes (i.e. water column nutrient content, nitracline, and effect of winter mixing and summer stratification on the shape of nitrate profile). The shape of the nitrate profile (i.e. correlation values), the temporal evolution of the 0-200 m averaged values and of the nitracline depth are consistent for the selected float in the Eastern Mediterranean Sea.

Figure 9.29 compares the seasonal time series of MLD, surface Chla, NO3, Si and PO4 in the North Atlantic during the "spring bloom", derived from the BGC-Argo floats observations and from the Global Ocean system of the Copernicus Marine Service. The percent bias and percent RMSD are also represented for each metric. The model reproduces the seasonal cycle of surface Chla and nutrients, i.e. the timings of minima, maxima, and the onset of the bloom, the winter Chla minimum and winter nutrients maxima. However, the skill metrics deteriorate in summer with the model underestimating Chla maximum and overestimating NO3 and PO4 minima.

The Global system skill for 22 metrics (Mignot et al., 2021) is summarised in the Taylor diagram (Figure 9.30), which allows for a rapid evaluation of strengths and weaknesses of a model simulation. For instance, the global model is skilled at reproducing oxygen levels, cycling of nutrients, and DIC, but the representation of Chla, POC, spCO2 and spH needs to be improved.

Finally, DM validation can be enriched by additional levels of process and system validation (Hipsey et al., 2020). These aim to assess the model performance, to simulate fluxes and rates of transformation, which drive changes in model state variables, and to verify emergent properties that are not directly predictable by the choices made to build the model structure and formulations. Measuring time and space variability of in-situ fluxes is difficult and highly resource consuming, thus the list of metrics remains restricted to few fluxes, such as rate of primary production, nutrient uptake, grazing rates, and sinking of organic particles. Nevertheless, the general confidence and fit-for-purpose in the underlying function of biogeochemical operational models can be increased by informing users about the uncertainty of a wider range of processes featured in the model formulation.

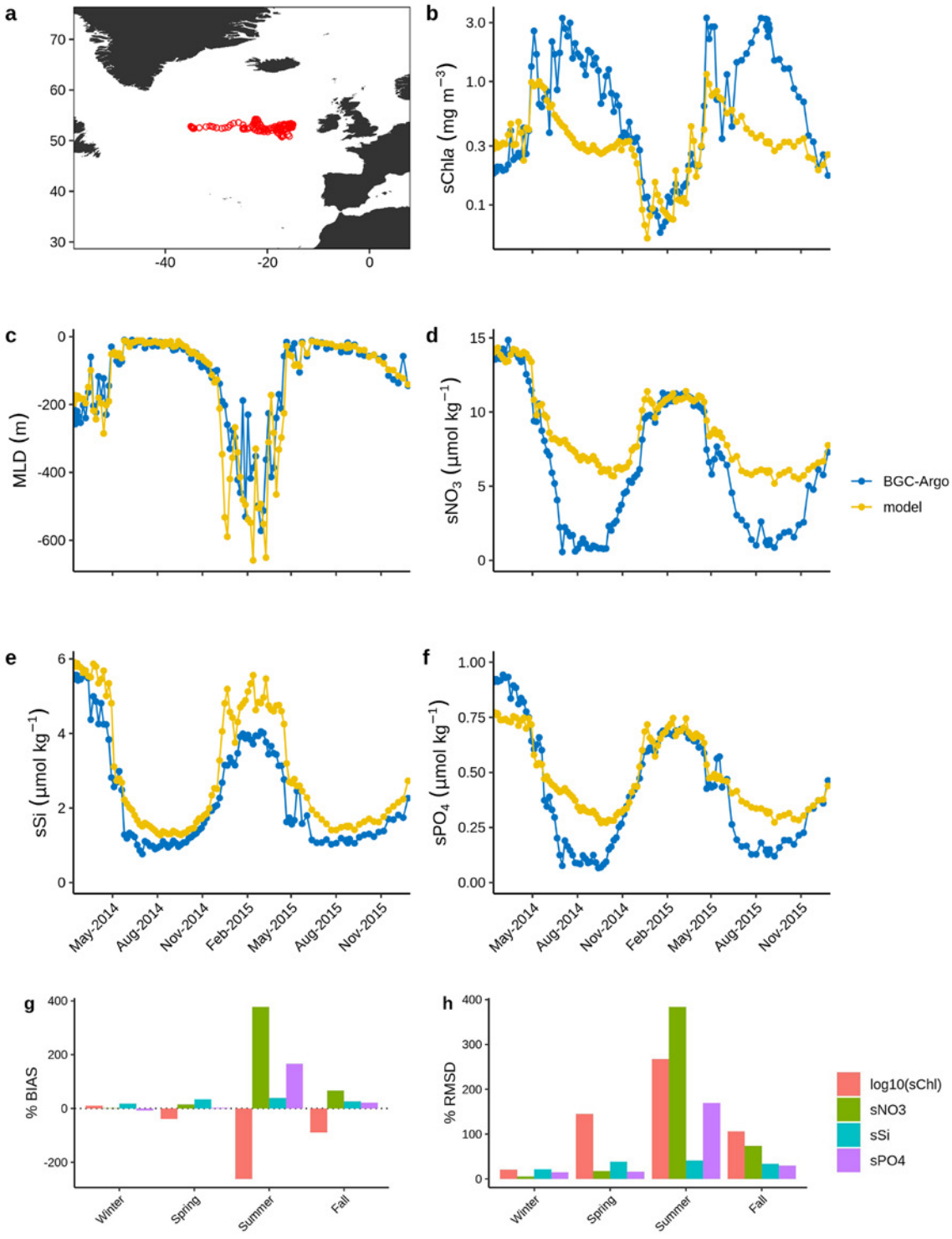


Figure 9.29. a): trajectory of a BGC-Argo float located in the North Atlantic. Time series derived from the BGC-Argo (blue) and the model simulation (yellow): b): mixed layer depth; c): surface Chla; d): NO_3; e): Si; f): PO_4. For each metric: g): seasonal percent bias; h): percent RMSD (from Mignot et al., 2021).

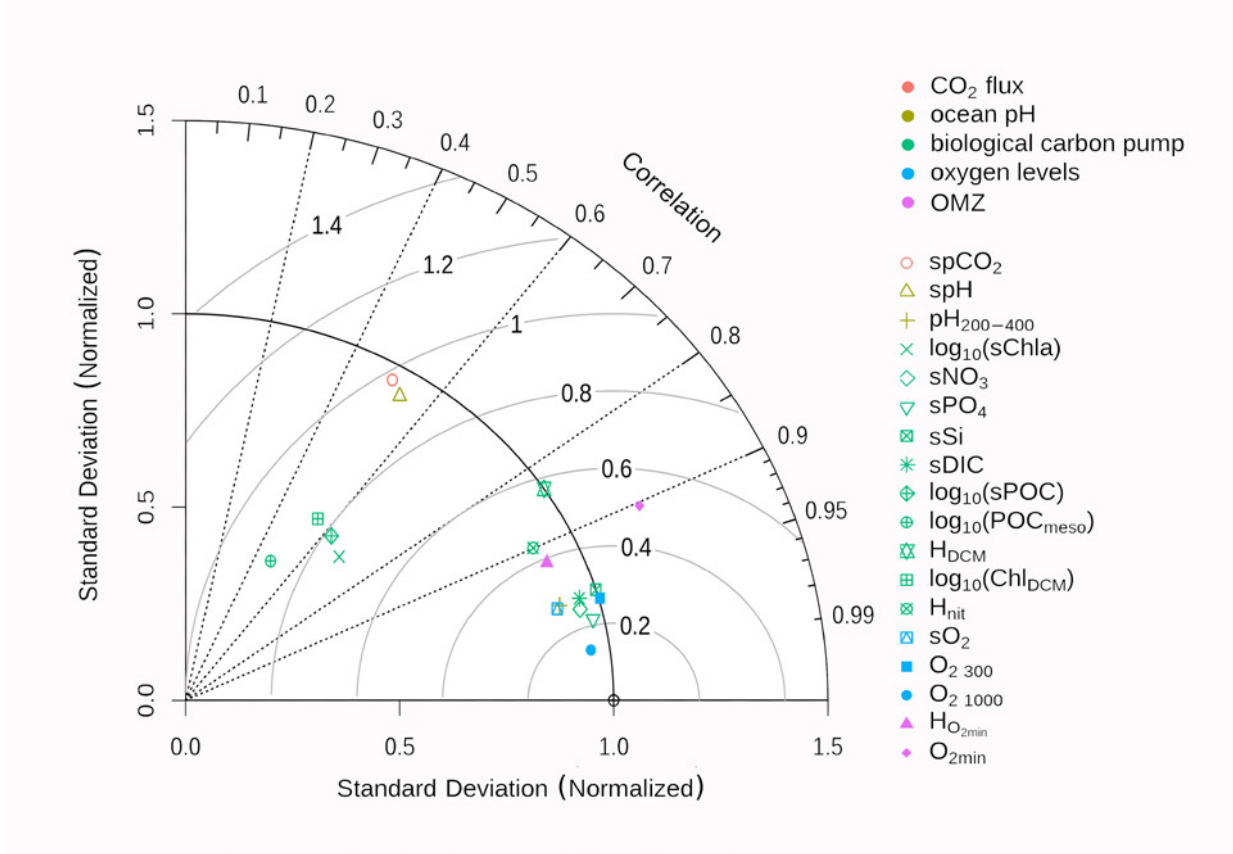


Figure 9.30. Comparison of BGC-Argo float observations and model values for 22 metrics using a Taylor diagram. The symbols correspond to the metrics and the colours represent the BGC processes with which they are associated (from Mignot et al., 2021).

9.2.7. Output

The purpose of this section is to provide recommendations and guidelines about the dissemination of products and the delivery of services based on BGC OOFS. These recommendations stem from the experience gained by some operational oceanography service centres, which deliver numerical products and have collected users' needs through the Service Desk, a structure dedicated to answer and manage any user request. Products and services, such as the production, preparation, and delivery of operational ocean forecasts to users in forms that meet their needs, are the final goal of an OOFS.

9.2.7.1. Data formats

Following the community of physical oceanographers, the biogeochemical community has widely adopted the NetCDF format (24) and the CF metadata conventions (25) for standard names and units. These standards are adopted by most operational oceanography actors (e.g. within GODAE OceanView), including the groups that operate numerical ocean prediction systems, and also by most of those delivering services based on oceanic observations.

24. https://www.unidata.ucar.edu/software/netcdf/
25. https://cfconventions.org/

9.2.7.2. Standard products

A BGC OOFS should offer users a reliable and easy access to valuable ocean information (past, present, and forecast). Each system operator should work to ensure that the following common variables (with their acronym or formula in brackets) are produced in delayed-mode and real time bases:

- nitrate concentration [NO_3]
- phosphate concentration [PO_4]
- dissolved oxygen concentration [O_2]
- chlorophyll-a concentration [Chla]
- phytoplankton concentration (expressed as carbon) [PHYC]
- net primary production of biomass (expressed as carbon) [NPP]

In addition to the above standard products, operators should also make available the following products, if they are represented in the model:

- silicate concentration [Si]
- iron concentration [Fe]
- ammonium concentration [NH_4]
- zooplankton concentration (expressed as carbon, mass, or mole) [ZOOC]
- PFTs chlorophyll-a concentration [PFTs]
- dissolved inorganic carbon concentration [DIC]
- total alkalinity [TALK]
- pH [pH]
- surface pCO_2 [$spCO_2$]
- air-sea CO_2 flux [fCO_2]
- light attenuation coefficient [Kd]
- photosynthetic photon flux [PAR]
- euphotic layer depth [ZEU]
- secchi_depth_of_sea_water [ZSD]

Model data are usually archived in the units specified by the International System of Units (SI Units), being mole per cubic metre (symbol mol m^{-3}) for concentration in seawater.

9.2.7.3. Data storage

The 2D or 3D concentrations of the modelled prognostics and diagnostics variables are saved and stored instantaneously, or averaged over specific time periods (daily, weekly, monthly, etc.). It has to be underlined that to store outputs requires substantial computer disk space, especially for biogeochemical models which can generate a lot of variables or derived quantities. This should be considered before the operational system is set up.

9.2.7.4. Other end-user products

Other data and information, called "end-user products", can be derived from or in addition to the standard products, with the purpose of building indicators of the marine environment for water quality monitoring, pollution control (eutrophication phenomena), food web indicators, etc.

9.2.7.5. Applications

The scientific community has identified key variables and indicators to evaluate current state and likely future conditions of the ocean, such as the EOV (from the GOOS Expert Panels) or the OMI (from the Copernicus Marine Service). Oxygen, chlorophyll-a, primary production, nutrients, pH, and CO2 air-sea flux are monitored to keep track of ocean health and changes, also to advise the policy makers. These indicators provide important information also for ecosystem-based fish management, sustainable aquaculture, and fisheries research. The number of users of BGC model products has been steadily increasing during the last years (Figure 9.31), highlighting the growing interest for BGC.

9.2.8. Higher trophic levels modelling

Researches by marine biologists, ecologists, and fishery scientists very often use a set of environmental variables to explain available observations for one species of interest and make predictions. Examples of frequently collected information include geo-referenced fishery catch data or acoustic-derived abundance of a fish species, scientific sampling of eggs, larvae or juveniles, satellite tracking of individuals of large fish, seabirds, turtles or marine mammals, or simply visual observations (whales). These studies are based on the correlation between

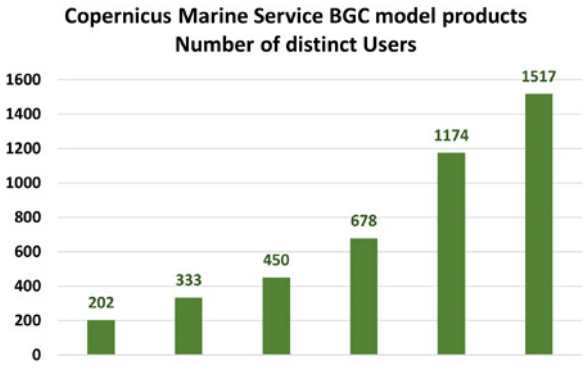


Figure 9.31. Number of distinct users of BGC model products of the Copernicus Marine Service during the last years (courtesy of the Service Desk).

outputs of the statistical or mechanistic model developed using the environmental variables and the observed variables, i.e. the presence or abundance of the studied species at a given stage of development. Since the relationships are defined using observations collected in a very dynamic environment with multiple sources of variability in time and space, it is essential to use environmental variables co-located with the observations. However, due to limitation in observations, marine scientists most often have to aggregate their data sets to crude resolutions, e.g. by season or year in large geographical boxes, or they restrain their analyses to satellite-derived oceanic variables, such as SST (available since early 1980s), SSH (since 1992), and sea surface Chla concentration (since 1998). The provision of these satellite-derived variables has generated large progress in the understanding of ecology and population dynamics of marine species. However, there are still some gaps in the use of these variables given that: i) satellites measure only the surface of the ocean; and ii) surface Chla is a proxy of primary production, which is not necessarily closely related to the upper trophic level animals that feed on zooplankton or larger organisms (e.g. micronekton). Furthermore, in the development phase of these organisms (lasting from days to months), the spatial and temporal correlation between primary production and these animals may be lost.

Modelling tools have the potential to fill these gaps, by simulating the marine food web with primary production, zooplankton, and micronekton as essential variables to support HTL. As explained in Section 9.1.1, BGC and HTL models are often separate models as they focus on different processes but BGC models can provide input for HTL models, and there are examples of BGC-HTL coupled models (e.g., Libralato and Solidoro, 2009; Rose et al., 2015; Aumont et al., 2018; Diaz et al., 2019). However, presently the link (online/offline coupling) is neither straightforward nor fully investigated. Thus, HTL models currently must rely also on other sources of input, such as satellite and in-situ data collection.

Connections, challenges, and expectations in bridging BGC and HTL modelling are discussed in the next subsections.

9.2.8.1. Essential variables

Primary production, zooplankton, and micronekton are essential ecosystem variables for the development of applications directed to management and conservation of marine resources and its biodiversity. Primary production is the source of energy to low and mid-trophic level functional groups. Zooplankton are a crucial link between the primary producers (mainly phytoplankton) and the micronekton at the mid-trophic level of the marine food web, as well as many mid-size pelagic species and some specialised large predators (e.g. baleen whales). Micronekton is defined by a size range between 1 and 10 cm, and include many species of fish, crustaceans and cephalopods, as well as the early life stages of many larger fish species. The micronekton that inhabit permanently the lower mesopelagic depths (~ below 300-400m) feed on the organic matter sinking in the water column. All micronekton organisms, including the species temporarily occupying this trophic level and size range, are the forage of larger marine species that have developed various skills to detect and feed on them.

Primary production, zooplankton, and micronekton are thus key inputs to investigate the mechanisms driving fish recruitment, as well as movement and migration of oceanic predators.

9.2.8.2. Satellite-derived and in-situ observations

9.2.8.2.1. Primary production

To establish which mechanisms control the distribution, recruitment, and abundance of large oceanic exploited or protected species, marine scientists require a three-dimensional representation of the environment and not only surface observations as those provided by satellites. The existence of a deep Chla maximum (e.g. in tropical waters and the Arctic) is a good illustration of the lack of adequation between surface and subsurface. One possible solution for this problem is to extrapolate the satellite observations over the water column according to some empirical models developed to estimate vertically integrated primary production, or NPP, based on surface Chla and key variables (SST and solar radiation). This product provides an essential foundation to monitor ocean productivity. However, various flaws remain, there are caveats for shallow waters and the Arctic, as well as difficulties in resolving persistently cloudy regions. However, primary production can also be provided by BGC models, offering the better three-dimensional vision as opposed to the satellite-based estimates, but this solution is still little used although the improvements in BGC models (in particular thanks to the use of data assimilation) are promising.

9.2.8.2.2. Zooplankton

Zooplankton is certainly the variable on which have been developed the most advanced applications on larval recruitment, fish habitat, dynamics of small and mid-size pelagic species as well as baleen whales. Despite decades of sampling efforts at sea, zooplankton observations remain limited to a few valuable long-term time series from oceanographic stations and a partial global climatology from the compilation of all available data collected, which represents a huge effort of data standardisation (Moriarty and O'Brien, 2013). Therefore, only numerical models can provide the synoptic maps of zooplankton distributions needed by ecologists and fishery scientists.

9.2.8.2.3. Micronekton

Micronekton species, including a huge biomass of mesopelagic organisms, are among the largest unknowns in the functioning of the global ocean ecosystem. This is a critical gap to understand the ecology of their predators for which there is a lot of interest in terms of resource management and conservation. In recent years, climate change, carbon storage in the deep ocean, and the role of diel vertical migration of mesopelagic (and zooplankton) have become major scientific issues.

But, even more than for zooplankton, the sparsity of observations on a global scale and over time poses a real problem for modellers of higher trophic levels. The traditional approach for sampling micronekton is net trawling. Many studies are simply qualitative descriptions of species, quite often used in combination with acoustic sampling to support the extrapolation of acoustic signal to biomass estimates. However, biomass based on acoustic sampling, especially with one single frequency, can be easily biased by one or two order of magnitudes due to the very strong resonance (backscatter) of some organisms, e.g. gelatinous organisms containing gas bubbles (Proud et al., 2018) or conversely very weak resonance despite large biomass, e.g. fish without swim bladder (Dornan et al. 2019; Escobar-Flores et al. 2019). In the absence of sufficient data coverage, relatively simple modelling approaches are used to simulate these functional groups in a food web model, relying on allometric scale relationships, first macro-ecological principles, or fluxes between trophic boxes.

9.2.8.3. Models of zooplankton and mid-trophic levels

9.2.8.3.1. Complexified BGC models

Improving resolution of primary production in BGC models helps to get better zooplankton predictions, although the relationship is not so straightforward. The reason is that in models the zooplankton component is used as the closure term of the biogeochemical cycles. To compensate for the lack in biogeochemical models of zooplankton predation by higher trophic levels, a mortality function with a mortality rate increasing rapidly (quadratic term) is used to avoid numerical instability at high levels of biomass. A high mortality rate is realistic in warm tropical waters but less for cold waters in which the lifespan of zooplankton is much longer, leading to high biomass persisting during fall. Underestimated zooplankton biomass can then have a cascading effect on the phytoplankton mortality. To address this issue in biogeochemical models, it may help the addition of a trophic level feeding on zooplankton, e.g. the micronekton at the intermediate trophic level, or a simplified representation of the entire upper food web with a size spectrum approach (Zhou et al 2010). Gelatinous organisms are also increasingly recognised as a key group in marine biogeochemical cycles as they need to be included to account for zooplankton mortality. A recent development consisted in the introduction of a jellyfish functional group in the biogeochemical model PLANKTOM (Wright et al., 2021), suggesting that it can have a large direct influence on the zooplankton as well as on the other groups through trophic cascades. However, parameterisation of physiological rates and validation of micronekton and jellyfish carbon biomass are limited by the deficit of data on these species. Moreover, adding mid-trophic level compartments would still increase complexity of BGC models, which are already complicated as including dozens of variables.

9.2.8.3.2. Spatially explicit models with transport

Models with less complexity and easier to parameterize can be used in the meantime. They are useful approaches complementary to more complex BGC models, allowing faster testing studies, e.g. for processes and new functional groups, with outputs providing useful intermediate benchmarks. These models do not include all the detailed biogeochemical cycles but focus on food web functional groups, size spectrum, or target species. The link with the lower trophic level can be as simple as an energy transfer coefficient between primary production and each functional group. A key advantage of reduced complexity is the greater facility to implement quantitative methods to estimate parameters using available observations, whether at global or regional level. Nevertheless, such models still simulate the transport by oceanic currents, either based on advection-diffusion equations like the ocean circulation and BGC models (Maury et al., 2007; Lehodey et al., 2010), or with mean transfers between adjacent grid cells and geographical boxes (Audzijonyte et al., 2019). Transport can be also simulated using Lagrangian IBM approaches that keep track of individuals or meta-individuals characterised by individual state variables (weight, length, energy storage, life stage, etc) and behavioural rules. However, due to computational cost, this approach is still usually limited to regional domains or single species (DeAngelis and Gross, 1992; Carlotti and Wolf, 1998; Miller et al., 1998; Huse et al., 2018).

9.2.8.3.3. Ecosystem food web and size spectrum models

Other modelling approaches are oriented towards a representation of the ecosystem food web to explore the interactions between fisheries and exploited, by-catch, or protected species (Christensen and Walters, 2004). Zooplankton and mid-trophic levels are often defined by a small number of functional groups in the food web interactions. The difficulty comes from the rapid increase in the number of parameters as functional groups and species are added in the food web model. The increasing complexity in the network of connections between the numerous groups, species and sometimes life stages of species, is developed at expense of the spatial description. A few approaches combine such complexity with

a semi-spatially explicit representation, i.e. through bulk-transfer between geographical regions or cells from various sizes (e.g. Audzijonyte et al., 2019). The size spectrum is an approach that strongly simplifies the view of a marine ecosystem.

The size-based ecosystem modelling is a classical approach that is used to predict biomass distribution and size-structure of marine consumers (see review in Blanchard et al., 2017). Jennings and Collingridge (2015) have developed this approach at global scale. The model predicts rates and magnitudes of energy flux from primary producers to consumers that depend on primary production, transfer efficiency, predator and prey body mass, and temperature. Biomass is estimated in the water column without considering the horizontal transport nor the vertical structure, and mesopelagic communities are not explicitly modelled. Maury et al. (2007) have developed a similar size-spectrum approach but that also accounts for the influence of spatial dynamics and vertical diel migration. Aumont et al. (2018) have fully coupled this latter model to a physical-biogeochemical model allowing to explore two-way interactions between lower and higher trophic levels of the pelagic ecosystem. Petrik et al. (2019 and 2020) have proposed another approach that discretizes the size spectrum into a few stage-structured functional groups as in De Roos et al. (2008). Their demographic system at each spatial grid cell is forced offline by vertically integrated temperature, vertically integrated zooplankton biomass concentrations and mortality losses, bottom temperature, and detrital fluxes, but there is no transport or fish movement.

There is no simple solution to model end-to-end ocean ecosystems (Fulton, 2010) but various approaches that reflect the different scientific questions that are investigated. The demand for greater details in taxonomic representation and population dynamics (including transport, recruitment, and migrations) of target species, creates major problems in calculation, estimation of parameters, and analysis of uncertainties, which may be a critical issue if the model has to support management and policy decisions. For these reasons, to formulate management advice for quotas of catches and/or effort and conservation measures, RFMOs mostly rely on standard stock assessment modelling approaches, fitted to key target species and fisheries. These models have been used since the 1960s and can integrate multiple sources of information to estimate the key parameters of population dynamics and fisheries for a single species (Maunder and Punt, 2013). However, they treat the environmental variability as noise that is removed from fishing data using standardisation methods or integrated as a random signal in the predicted recruitment process, and thus they cannot be used to project mid- to long-term changes (e.g. climate change effects on fisheries).

9.2.8.4. Contribution from operational oceanography

Improved BGC models with assimilation of in-situ and satellite data is an approach with promising results and rapid progress. Thanks to data assimilation, the physical and biogeochemical models used in operational oceanography to predict and forecast ocean physics and primary production are becoming more and more accurate. Consequently, they are used by an increasing number of marine biologists, ecologists, and fishery scientists. The outputs of biogeochemical models are also essential to explore the historical period before the satellite era, which started in the late 1970s (see Figure 9.7). The information generated by BGC models is also needed to develop seasonal forecasting of ocean ecosystems, population dynamics of marine animals, and to explore the impact of climate change with long-term projections, once forced by Earth System Models. Many BGC models also provide dissolved oxygen concentration and pH, which are useful variables for modelling habitats of fishes. Finally, the recent progress achieved in operational oceanography contributes to an overall improvement of all types of zooplankton, micronekton and ecosystem models.

A global zooplankton and micronekton model-based product (Lehodey et al., 2010 and 2015) is delivered in the Copernicus Marine Service. With only 11 parameters, the model simulates one functional group of zooplankton and six functional groups of micronekton in the global ocean, with a vertical structure simplified into three layers in the water column (epipelagic, and upper- and lower-mesopelagic) allowing to consider vertically migrant and non-migrant mesopelagic behaviours. The functional groups are driven by primary production, euphotic depth, temperature, and horizontal currents with time of development and mortality rate linked to water temperature. The limited number of parameters allows implementing quantitative methods to estimate their optimal values by searching for the best fit between observations and predictions (Lehodey et al., 2015). However, the sparsity of direct biomass observations and the difficulty to convert the signal of acoustic echo-sounders into biomass is still an issue that requires further developments. In particular, there is the need to progress on acoustic models (Jech et al., 2015).

9.2.8.5. Applications

Zooplankton and micronekton outputs produced by the Copernicus Marine Service have proved to be useful variables along with physical and biogeochemical variables to model feeding habitats, feeding behaviour, and migrations of large oceanic protected species such as marine mammals and turtles (e.g., Abecassis et al., 2013; Lambert et al., 2014; Chambault et al., 2016; Roberts et al., 2016; Green et al., 2020; Pérez-Jorge et al., 2020; Romagosa et al., 2020 and 2021). These applications contribute to the scientific advice needed to propose marine spatial management measures (e.g., Ma-

rine Protected Areas and Migratory Corridors), the planning of activities at seas (e.g., offshore energy, military exercises and tests, and navigation routes), and real-time operational tools to limit the interaction of fisheries with protected species (Howell et al., 2008; Hobday et al., 2010; Hazen et al. 2018). The combination of zooplankton and micronekton variables has been used in a mechanistic model of Antarctic krill population, including food conditions that adults need to successfully produce eggs and the density of predators feeding on spawned eggs (Green et al. 2021).

Finally, spatially explicit population dynamics of target species can be driven by these variables to study recruitment, natural mortality, and movements linked to feeding behaviour and spawning migrations of fish (Lehodey et al., 2008; Dueri et al., 2012; Hernandez et al., 2014; Scutt Phillips et al., 2018; Senina et al., 2019). These models, combined with quantitative methods integrating various sources of georeferenced data (i.e. catch, size frequencies of catch, tagging data, density of larvae, and acoustic biomass abundance), provide new tools to assess the status of exploited stocks (Senina et al., 2008 and 2020; Dragon et al., 2018), to test spatial management scenarios (Sibert et al., 2012), to develop real time monitoring applications (Lehodey et al., 2017), and forecast seasonal to long-term changes along with IPCC climate scenarios (Lehodey et al., 2013; Dueri et al., 2014; Bell et al., 2013 and 2021).

9.2.9. Inventories

The first Green Ocean applications of operational oceanography, coupling biogeochemical models, and assimilation components from the existing GODAE systems, were discussed in Brasseur et al. (2009). Some years later, Gehlen et al. (2015) and Fennel et al. (2019) discussed the current state and future prospects of analysis and prediction tools for ocean biogeochemistry and ecosystems, and presented representative examples of global and regional physical–biogeochemical systems implemented in pre-operational or operational mode. Currently, a few forecasting systems are fully operational, i.e. maintained by an operational centre with strict commitment to routinely provide forecasts.

Tables 9.1 and 9.2 provide initial inventories of the operational forecasting and multi-year systems, based on the literature mentioned above and completed in collaboration with the MEAP-TT working group that is one of the OceanPredict Task Teams. General information is given for each system, along with type (from global to coastal scale), producer, resolution, implemented model, data assimilation method, and product catalogue, as well as the web address that the reader can consult for further details.

Table 9.1. Initial inventory of BGC Global (G) to Regional (R) to Coastal (C) operational forecasting systems.

Type	System	Covered Area	Resolution	PHY-BGC models	BGC Data Assimilation (method and data)	Products	Website
G	Global Ocean BGC system (MOI, France)	Global ocean	1/4°	PISCES coupled offline with NEMO (1/12° degraded to 1/4°) at daily frequency	SEEK method, using total Chla from OC satellite data	Chla, NO3, PO4, Si, Fe, O2, PHYC, NPP, spCO2, pH, 10-days forecast, updated weekly	https://marine.copernicus.eu

Type	System	Covered Area	Resolution	PHY-BGC models	BGC Data Assimilation (method and data)	Products	Website
R	Northwest European Shelf Seas BGC system (UK Metoffice, UK)	European North-West shelf Seas	~7 km	ERSEM coupled online with NEMO	3D-Var NEMOVAR method, using total Chla from OC satellite data	Chla, NO3, PO4, O2, PHYC, NPP, spCO2, pH, Kd, 6-day forecast, updated daily	https://marine.copernicus.eu
R	TOPAZ5-ECOSMO Arctic Ocean system (Norwegian Meteorological Institute, Norway; Nansen Environmental and Remote Sensing Center, Norway)	Arctic Region	6 km	ECOSMO biological model coupled online to the HYCOM ocean physical model	Assimilates Chla from OC satellite data using a nudging approach, and surface observations are projected downward in the water column applying an algorithm described by Uitz et al. (2006).	Chla, NO3, PO4, Si, O2, PHYC, ZOOC, NPP, spCO2, DIC, pH, Kd, 10-day forecast, updated daily	https://marine.copernicus.eu
R	Baltic Sea system (Swedish Meteorological and Hydrological Institute, Sweden)	Baltic Sea	1 nautical mile	ERGOM coupled online with NEMO	_	Chla, NO3, PO4, NH4, O2, spCO2, pH, NPP, ZSD, 6-day forecast, updated twice daily	https://marine.copernicus.eu
R	Iberia-Biscay-Irish system (MOI, France + consortium)	Iberian-Biscay-Irish shelves	1/36°	NEMO-PISCES online coupled model; nested into PHY and BGC solutions from the Global MFC	No assimilation	Chla, NO3, NH4, PO4, Si, Fe, O2, PHYC, NPP, spCO2, DIC, pH, ZEU, 10-days forecast updated on a weekly basis	https://marine.copernicus.eu
R	MedBFM3 model system (Euro Mediterranean Center on Climate Change - CMCC, Italy; OGS, Italy)	Mediterranean Sea	1/24°	BFM v5 model, off-line coupled with NEMO	3DVAR-BIO method, using Chla from satellite and vertical profiles of Chla and nitrate from BGC-Argo	Chla, PHYC, ZOOC, NO3, NH4, PO4, Si, O2, spCO2, pH, fCO2, ALK, DIC, NPP, 10-day forecast updated daily	https://marine.copernicus.eu

Type	System	Covered Area	Resolution	PHY-BGC models	BGC Data Assimilation (method and data)	Products	Website
R	Black Sea system (University of Liege, Belgium)	Black Sea	~3km	BAMHBI, online coupled with NEMO	"Ocean Assimilation Kit" (OAK; Vandenbulcke and Barth, 2015) for assimilation of surface Chla from satellite	Chla, PHYC, NO3, PO4, Si, NH4, O2, spCO2, pH, fCO2, ALK, DIC, NPP, Kd, PAR, 10-day forecast produced daily	https://marine.copernicus.eu
R	POSEIDON system (HCMR, Greece)	Mediterranean Sea	1/10°	ERSEM-II model, on-line coupled with POM	No assimilation	Chla, PHYC, ZOOC, BACC, NO3, NH4, PO4, 4-day forecast updated daily	https://www.poseidon.hcmr.gr
C	J-SCOPE forecast system (JISAO's Seasonal Coastal Ocean Prediction of the Ecosystem, funded by NOAA, US)	California Current System	1/10°	ROMS ocean model coupled with a BGC model	_	Seasonal forecasts of sea surface temperature (SST) and BGC variables	http://www.nanoos.org/products/j-scope/home.php
C	Harmful Algal Bloom Monitoring System (National Centers for Coastal Ocean Science, formed by the NOAA, US)	Coastal and lake regions of the US	_	_	_	Daily forecast	https://coastalscience.noaa.gov/research/stressor-impacts-mitigation/hab-monitoring-system/
C	Great Barrier Reef (Bureau of Meteorology et al.)	Great Barrier Reef	_	CSIRO eReefs modeling suite	_	A few days forecast	https://ereefs.org.au/ereefs
C	Chesapeake Bay	Chesapeake Bay	600m	ChesROMS-ECB	_	Nowcasts and a few days forecasts of physical and BGC variables (focusing on O2, acidification metrics, T, S)	www.vims.edu/hypoxia; https://oceansmap.maracoos.org/chesapeake-bay/

Table 9.2. Inventory of BGC Global (G) to Regional (R) to Coastal (C) multi-year systems.

Type	System	Covered Area	Resolution	PHY-BGC models	BGC Data Assimilation (method and data)	Products	Website
G	Global Ocean BGC system (MOI, France)	Global ocean	1/4°	PISCES, coupled offline with NEMO at daily frequency	No assimilation	Chla, NO3, PO4, Si, Fe, O2, PHYC, NPP, spCO2, pH, 1993 onwards	https://marine.copernicus.eu
G	Global Ocean low and mid-trophic levels product (CLS, France)	Global ocean	1/12°	LMTL component of SEAPODYM dynamical population model, driven offline by NEMO, NPP from satellite and PISCES	No assimilation	2D fields of zooplankton biomass and six groups of micronekton biomass, 1998 onwards	https://www.cls.fr/
R	Northwest European Shelf Seas BGC system (UK Met Office, UK)	European North-West shelf Seas	~7 km	ERSEM, coupled online with NEMO	3D-Var NEMOVAR method, using surface PFT Chla from OC satellite data	Chla, PFTs, PHYC, NO3, PO4, O2, spCO2, pH, NPP, Kd, 1993 onwards	https://marine.copernicus.eu
R	TOPAZ-ECOSMO reanalysis system (Nansen Environmental and Remote Sensing Center, Norway)	Arctic Region	25 km	ECOSMO biological model coupled online to the HYCOM ocean physical model	Assimilates surface Chla a from OC satellite and in-situ nutrient profiles, using an Ensemble Kalman Smoother (EnKS) method, after a gaussian anamorphosis for all BGC data. EnKS is preferred to EnKF in delayed mode	Chla, NO3, PO4, O2, PHYC, ZOOC, Kd, 2007 onwards	https://marine.copernicus.eu
R	Baltic Sea system (Swedish Meteorological and Hydrological Institute, Sweden)	Baltic Sea	1 nautical mile	SCOBI coupled to NEMO	LSEIK data assimilation scheme, using oxygen and nutrients	Chla, NO3, NH4, PO4, O2, 1993 onwards	https://marine.copernicus.eu

Type	System	Covered Area	Resolution	PHY-BGC models	BGC Data Assimilation (method and data)	Products	Website
R	Iberia Biscay Irish system (MOI, France)	Irish-Biscay-Iberian shelves	1/12°	NEMO-PISCES online coupled model; nested into PHY and BGC solutions from the Global MFC	No assimilation	Chla, NO3, NH4, PO4, Si, Fe, O2, PHYC, NPP, spCO2, DIC, pH, ZEU, 1993 onwards	https://marine.copernicus.eu
R	Global Ocean low and mid-trophic levels product (CLS, France)	Global ocean	1/12°	LMTL component of SEAPODYM dynamical population model, driven offline by NEMO, NPP from satellite and PISCES	No assimilation	2D fields of zooplankton biomass and six groups of micronekton biomass, 1998 onwards	https://marine.copernicus.eu
R	MedBFM3 model system (OGS, Italy)	Mediterranean Sea	1/24°	BFM v5 model, off-line coupled with NEMO	3DVAR-BIO method, using surface Chla	Chla, PHYC, ZOOC, NO3, NH4, PO4, Si, O2, spCO2, pH, fCO2, ALK, DIC, NPP, 1999 onwards	https://marine.copernicus.eu
R	Black Sea system (University of Liege, Belgium)	Black Sea	~3km	BAMHBI model, online coupled with NEMO	No assimilation	Chla, PHYC, O2, NO3, PO4, spCO2, pH, fCO2, ALK, DIC, NPP, 1992 onwards	https://marine.copernicus.eu
R	China Sea Multi-Scale Ocean Modelling System (CMOMS)	China Seas	~3km	ROMS ocean model coupled with a BGC model	No assimilation	Chla, PHYC, ZOOC, NO3, NH4, PO4, O2, spCO2, pH, ALK, DIC, small detritus, large detritus, terrestrial POM, and terrestrial DOM; 1992 onwards	https://ocean.ust.hk:8443/SiteMapApi/new/index.jsp

9.3. References

Abecassis, M., Senina, I., Lehodey, P., Gaspar, P., Parker, D., Balazs, G., Polovina, J. (2013). A Model of Loggerhead Sea Turtle (Caretta caretta) Habitat and Movement in the Oceanic North Pacific. *PLoS ONE*, 8(9), e73274, https://doi.org/10.1371/journal.pone.0073274

Allen, J. I., M. Eknes, and G. Evensen. (2003). An Ensemble Kalman Filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea. *Annales Geophysicae*, 21(1), 399-411, https://doi.org/10.5194/angeo-21-399-2003

Almroth-Rosell, E., Eilola, K., Kuznetsov, Hall, I.P.O.J., and Meier, H.E.M. (2015). A new approach to model oxygen dependent benthic phosphate fluxes in the Baltic Sea. *Journal of Marine Systems*, 144, 127-141, https://doi.org/10.1016/j.jmarsys.2014.11.007

Anderson, L.A., Robinson, A.R., Lozano, C.J. (2000). Physical and biological modeling in the Gulf Stream region: I. Data assimilation methodology. *Deep Sea Research Part I: Oceanographic Research Papers*, 47, 1787-1827, https://doi.org/10.1016/S0967-0637(00)00019-4

Audzijonyte, A., Pethybridge, H., Porobic, J., Gorton, R., Kaplan, I., Fulton, E.A. (2019). Atlantis: A spatially explicit end-to-end marine ecosystem model with dynamically integrated physics, ecology and socio-economic modules. *Methods in Ecology and Evolution*, 10, 1814-1819, https://doi.org/10.1111/2041-210X.13272

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M. (2015). PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, 8, 2465-2513, https://doi.org/10.5194/gmd-8-2465-2015

Aumont O., Maury O., Lefort S., Bopp L. (2018). Evaluating the Potential Impacts of the Diurnal Vertical Migration by Marine Organisms on Marine Biogeochemistry. *Global Biogeochemical cycles*, 32 (11), 1622-1643, https://doi.org/10.1029/2018GB005886

Baretta, J. W., Ebenhöh, W., and Ruardij, P. (1995). The European regional seas ecosystem model, a complex marine ecosystem model. *Netherlands Journal of Sea Research*, 33, 233-246, https://doi.org/10.1016/0077-7579(95)90047-0

Bell, J.D., Ganachaud, A., Gehrke, P.C., Griffiths, S.P., Hobday, A.J., Hoegh-Guldberg, O., Johnson, J.E., Le Borgne, R., Lehodey, P., Lough, J.M., Matear, R.J., Pickering, T.D., Pratchett, M.S., Sen Gupta, A., Senina, I. and Waycott, M. (2013). Mixed response of tropical Pacific fisheries and aquaculture will respond differently to climate change. Nature Climate Change, 3, 591-599, https://doi.org/10.1038/nclimate1838

Bell, J.D., Senina, I., Adams, T., Aumont, O., Calmettes, B., Clark, S., Dessert, M., Hampton, J., Hanich, Q., Harden-Davies, H., Gehlen, M., Gorgues, T., Holmes, G., Lehodey, P., Lengaigne, M., Mansfield, B., Menkes, C., Nicol, S., Pasisi, C., Pilling, G., Ota, Y., Reid, C., Ronneberg, E., Sen Gupta, A., Seto, K., Smith, N., Taei, S., Tsamenyi, M., Williams, P. (2021). Pathways to sustaining tuna-dependent Pacific Island economies during climate change. *Nature Sustainability*, 4, 900-910, https://doi.org/10.1038/s41893-021-00745-z

Berline, L., Brankart, J.M., Brasseur, P., Ourmières, Y., and Verron, J. (2007). Improving the physics of a coupled physical–biogeochemical model of the North Atlantic through data assimilation: Impact on the ecosystem. *Journal of Marine Systems*, 64(1-4), 153-172, https://doi.org/10.1016/j.jmarsys.2006.03.007

Berthet, S., Séférian, R., Bricaud, C.,Chevallier, M., Voldoire, A., and Ethé, C. (2019). Evaluation of an online grid-coarsening algorithm in a global eddy-admitting ocean biogeochemical model. *Journal of Advances in Modeling Earth Systems*, 11, 1759-1783, https://doi.org/10.1029/2019MS001644

Bertino, L., Evensen, G., and Wackernagel, H. (2003). Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71(2), 223-241, https://doi.org/10.1111/j.1751-5823.2003.tb00194.x

Biogeochemical-Argo Planning Group (2016). The Scientific Rationale, Design andImplementation Plan for a Biogeochemical-Argo Float Array. https://doi.org/10.13155/46601

Blanchard, J.L., Heneghan, R.F., Everett, J.D., Trebilco, R., Richardson, A.J. (2017). From bacteria to whales: Using function size spectra to model marine ecosystems. *Trends in Ecology and Evolution*, 32(3), 174-186, https://doi.org/10.1016/j.tree.2016.12.003

Boyer, T.P., Baranova, O.K., Coleman, C., Garcia, H.E., Grodsky, A., Locarnini, R.A., Mishonov, A.V., Paver, C.R., Reagan, J.R., Seidov, D., Smolyar, I.V., Weathers, K., Zweng, M.M. (2018). World Ocean Database 2018. A.V. Mishonov, Technical Ed., NOAA Atlas NESDIS 87, https://www.ncei.noaa.gov/products/world-ocean-database

Brankart, J.-M., Testut, C.-E., Béal, D., Doron, M., Fontana, C., Meinvielle, M., and Brasseur, P. (2012). Towards an improved description of oceanographic uncertainties : effect of local anamorphic transformations on spatial correlations. *Ocean Science*, 8, 121-142, https://doi.org/10.5194/os-8-121-2012

Brasseur, P., Gruber, N., Barciela, R., Brander, K., Doron, M., El Moussaoui, A., Hobday, M. Huret, A.J., Kremeur, A.-S., Lehodey, P., and others. (2009). Integrating biogeochemistry and ecology into ocean data assimilation systems. *Oceanography*, 22(3), 206-215, https://doi.org/10.5670/oceanog.2009.80

Bricaud, C., Le Sommer, J., Madec, G., Calone, C., Deshayes, J., Éthé, C., Chanut, J., Lévy, M. (2020). Multigrid algorithm for passive tracer transport in the NEMO ocean circulation model: a case study with the NEMO OGCM (version 3.6). *Geoscientific Model Development*, 13(11), 5465-5483, https://doi.org/10.5194/gmd-13-5465-2020

Brown, C. D., and Davis, H. T. (2006). Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80(1), 24-38 https://doi.org/10.1016/j.chemolab.2005.05.004

Butenschön, M., Clark, J., Aldridge, J. N., Allen, J. I., Artioli, Y., Blackford, J., Bruggeman, J., Cazenave, P., Ciavatta, S., Kay, S., Lessin, G., van Leeuwen, S., van der Molen, J., de Mora, L., Polimene, L., Sailley, S., Stephens, N., and Torres, R. (2016). ERSEM 15.06: a generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels. *Geoscientific Model Development*, 9, 1293-1339, https://doi.org/10.5194/gmd-9-1293-2016

Campbell, J. (1995). The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research: Oceans*, 100, C7, 13237-13254, https://doi.org/10.1029/95JC00458

Capet, A., Meysman, F. J. R., Akoumianaki, I., Soetaert, K., and Grégoire, M. (2016). Integrating sediment biogeochemistry into 3-D oceanic models: A study of benthic-pelagic coupling in the Black Sea. *Ocean Modelling*, 101, 83-100, https://doi.org/10.1016/j.ocemod.2016.03.006

Carlotti, F., Wolf, K.U. (1998). A Lagrangian ensemble model of Calanus finmarchicus coupled with a 1-D ecosystem model. *Fisheries Oceanography*, 7, 191-204, https://doi.org/10.1046/j.1365-2419.1998.00085.x

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, e535, https://doi.org/10.1002/wcc.535

Carroll, D., Menemenlis, D., Adkins, J.F., Bowman, K.W., Brix, H., Dutkiewicz, S., Fenty, I., Gierach, M.M., Hill, C., Jahn, O., and Landschützer, P. (2020). The ECCO-Darwin data-assimilative global ocean biogeochemistry model: Estimates of seasonal to multidecadal surface ocean pCO2 and air-sea CO2 flux. *Journal of Advances in Modeling Earth Systems*, 7(3-4), 191-204, https://doi.org/10.1029/2019MS001888

Chai, F., Johnson, K., Claustre, H., Xing, X., Wang, Y., Boss, E., Riser, S., Fennel, K., Schofield, O., Sutton, A. (2020). Monitoring ocean biogeochemistry with autonomous platforms. *Nature Reviews Earth & Environment*, 1, 315-326, https://doi.org/10.1038/s43017-020-0053-y

Chambault P., de Thoisy B., Heerah K., Conchon A., Barrioz S., Dos Reis V., Berzins R., Kelle L., Picard B., Roquet F., Le Maho Y., Chevallier D. (2016). The influence of oceanographic features on the foraging behavior of the olive ridley sea turtle Lepidochelys olivacea along the Guiana coast. *Progress in Oceanography*, 142: 58-71, https://doi.org/10.1016/j.pocean.2016.01.006

Christensen, V., and Walters, C. (2004). Ecopath With Ecosim: Methods, Capabilities and Limitations. *Ecological Modelling*, 172, 109-139, https://doi.org/10.1016/j.ecolmodel.2003.09.003

Ciavatta, S., Torres, R., Saux-Picart, S. and Allen, J.I. (2011). Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? *Journal of Geophysical Research: Oceans*, 116(C12), https://doi.org/10.1029/2011JC007219

Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., Dall'Olmo, G., Polimene, L., and Allen, J. I. (2014). Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. *Progress in Oceanography*, 127, 74-95, https://doi.org/10.1016/j.pocean.2014.06.002

Ciavatta, S., Kay, S., Saux-Picart, S., Butenschön, M. and Allen, J.I. (2016). Decadal reanalysis of biogeochemical indicators and fluxes in the North West European shelf-sea ecosystem. *Journal of Geophysical Research: Oceans*, 121(3), 1824-1845, https://doi.org/10.1002/2015JC011496

Ciavatta, S., Brewin, R. J. W., Skákala, J., Polimene, L., de Mora, L., Artioli, Y., and Allen, J. I. (2018). Assimilation of ocean-color plankton functional types to improve marine ecosystem simulations. *Journal of Geophysical Research: Oceans*, 123, 834-854, https://doi.org/10.1002/2017JC013490

Ciavatta, S., Kay, S., Brewin, R.J., Cox, R., Di Cicco, A., Nencioli, F., Polimene, L., Sammartino, M., Santoleri, R., Skakala, J. and Tsapakis, M. (2019). Ecoregions in the Mediterranean Sea through the reanalysis of phytoplankton functional types and carbon fluxes. *Journal of Geophysical Research: Oceans*, 124(10), 6737-6759, https://doi.org/10.1029/2019JC015128

Cossarini, G., Mariotti, L., Feudale, L., Mignot, A., Salon, S., Taillandier, V., Teruzzi, A., D'Ortenzio, F. (2019). Towards operational 3D-Var assimilation of chlorophyll Biogeochemical-Argo float data into a biogeochemical model of the Mediterranean Sea. *Ocean Modelling*, 133, 112-128, https://doi.org/10.1016/j.ocemod.2018.11.005

Daewel, U., and Schrum, C. (2013). Simulating long-term dynamics of the coupled North Sea and Baltic Sea ecosystem with ECOSMO II: Model description and validation. *Journal of Marine Systems*, 119-120, 30-49, https://doi.org/10.1016/j.jmarsys.2013.03.008

Dall'Olmo, G., Dingle, J., Polimene, L., Brewin, R.J.W., Claustre, H. (2016). Substantial energy input to the mesopelagic ecosystem from the seasonal mixed-layer pump. *Nature Geoscience*, 9, 820-823, https://doi.org/10.1038/ngeo2818

De Roos, A.M., Schellekens, T., Van Kooten, T., Van De Wolfshaar, K., Claessen, D., Persson, L. (2008). Simplifying a physiologically structured population model to a stage-structured biomass model. *Theoretical Population Biology*, 73, 47-62, https://doi.org/10.1016/j.tpb.2007.09.004

DeAngelis, D.L., Gross, L.J. (1992). Individual-based Models and Approaches in Ecology. Chapman & Hall, London.

Diaz, F., D Bănaru, P. Verley and Y.-J. Shin (2019). Implementation of an end-to-end model of the Gulf of Lions ecosystem (NW Mediterranean Sea). II. Investigating the effects of high trophic levels on nutrients and plankton dynamics and associated feedbacks. *Ecological Modelling*, 405, 51-68, https://doi.org/10.1016/j.ecolmodel.2019.05.004

Dickey, T.D. (1991). The emergence of concurrent high-resolution physical and bio-optical measurements in the upper ocean and their applications. *Reviews on Geophysics*, 29(3), 383- 413, https://doi.org/10.1029/91RG00578

Dornan, T., Fielding, S., Saunders, R.A., Genner, M.J. (2019). Swimbladder morphology masks Southern Ocean mesopelagic fish biomass. *Proceedings of the Royal Society B Biological Science*, 286(1903), http://dx.doi.org/10.1098/rspb.2019.0353

Doron, M., Brasseur, P., Brankart, J.M. (2011). Stochastic estimation of biogeochemical parameters of a 3d ocean coupled physical-biogeochemical model: twin experiments. *Journal of Marine Systems*, 87, 194-207, https://doi.org/10.1016/j.jmarsys.2013.02.007

Dragon A-C., Senina, Hintzen N.T., Lehodey P. (2018). Modelling South Pacific Jack Mackerel spatial population dynamics and fisheries. *Fisheries Oceanography*, 27(2), 97-113, https://doi.org/10.1111/fog.12234

Dueri S., Faugeras B., Maury O. (2012). Modelling the skipjack tuna dynamics in the Indian Ocean with APECOSM-E: Part 1. Model formulation. *Ecological Modelling*, 245, 41-54, https://doi.org/10.1016/j.ecolmodel.2012.02.008

Dueri, S., Bopp, L., Maury, O. (2014). Projecting the impacts of climate change on skipjack tuna abundance and spatial distribution. *Global Change Biology*, 20(3): 742-753, https://doi.org/10.1111/gcb.12460

Dupont, F. (2012). Impact of sea-ice biology on overall primary production in a biophysical model of the pan-Arctic Ocean. Journal of Geophysical Research: Oceans, 117(C8), https://doi.org/10.1029/2011JC006983

Eilola, K., Meier, H.E.M., Almroth, E. (2009). On the dynamics of oxygen, phosphorus and cyanobacteria in the baltic sea: A model study. *Journal of Marine Systems*, 75, 163-184, https://doi.org/10.1016/j.jmarsys.2008.08.009

Escobar-Flores, P.C., O'Driscoll, R.L., Montgomery, J.C. (2018). Spatial and temporal distribution patterns of acoustic backscatter in the New Zealand sector of the Southern Ocean. *Marine Ecology Progress Series*, 592, 19-35, https://doi.org/10.3354/meps12489

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5), 10143-10162, https://doi.org/10.1029/94JC00572

Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C. A., Ford, D., Friedrichs, M. A. M., Grégoire, M., Jones, E., Kim, H.-C., Lamouroux, J., Murtugudde, R., Perruche, C. and the GODAE OceanView Marine Ecosystem Analysis and Prediction Task Team (2019). Advancing marine biogeochemical and ecosystem reanalyses and forecasts as tools for monitoring and managing ecosystem health. Frontiers in Marine Science, 6, 89, https://doi.org/10.3389/fmars.2019.00089

Feudale, L., Teruzzi, A., Salon, S., Bolzon, G., Lazzari, P., Coidessa, G., Di Biagio, V., Cossarini, G. (2021). Product Quality Document For the Mediterranean Sea Production Centre, MEDSEA_ANALYSISFORECAST_BGC_006_014, https://doi.org/10.25423/cmcc/medsea_analysisforecast_bgc_006_014_medbfm3

Flynn, K. J., Stoecker, D. K., Mitra, A., Raven, J. A., Glibert, P. M. Hansen, P. J., Granéli, E., and Burkholder, J. M. (2013). Misuse of the phytoplankton-zooplankton dichotomy: the need to assign organisms as mixotrophs within plankton functional types. *Journal of Plankton Research*, 35, 3-11, https://doi.org/10.1093/plankt/fbs062

Follows, M. J., Dutkiewicz, S., Grant, S., and Chisholm, S. W. (2007). Emergent biogeography of microbial communities in a model ocean. *Science*, 315(5820), 1843-1846. https://doi.org/10.1126/science.1138544

Fontana, C., Brasseur, P., and Brankart, J. M. (2013). Toward a multivariate reanalysis of the North Atlantic Ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data. *Ocean Science*, 9(1), 37-56, https://doi.org/10.5194/os-9-37-2013

Ford, D., and Barciela, R. (2017). Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products. *Remote Sensing of Environment*, 203, 40-54, https://doi.org/10.1016/j.rse.2017.03.040

Ford, D., Key, S., McEwan, R., Totterdell, I. and Gehlen, M. (2018). Marine biogeochemical modelling and data assimilation for operational forecasting, reanalysis, and climate research. *New Frontiers in Operational Oceanography*, 625-652, https://doi.org/10.17125/gov2018.ch22

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, https://doi.org/10.5194/bg-18-509-2021

Fulton, E.A. (2010). Approaches to end-to-end ecosystem models. *Journal of Marine Systems* 81, 171-183, https://doi.org/10.1016/j.jmarsys.2009.12.012

Fulton, E., Smith, A., Johnson, C. (2003). Effect of Complexity of Marine Ecosystem Models. *Marine Ecology Progress Series*, 253:1-16, doi:10.3354/meps253001

Garcia, H.E., Weathers, K., Paver, C.R., Smolyar, I., Boyer, T.P., Locarnini, R.A., Zweng, M.M, Mishonov, A.V., Baranova, O.K., Seidov, D., and Reagan, J.R. (2018a). World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. A. Mishonov Technical Ed.; NOAA Atlas NESDIS 83, 38pp.

Garcia, H.E., Weathers, K., Paver, C.R., Smolyar, I., Boyer, T.P., Locarnini, R.A., Zweng, M.M, Mishonov, A.V., Baranova, O.K., Seidov, D., and Reagan, J.R. (2018b). World Ocean Atlas 2018, Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate). A. Mishonov Technical Ed.; NOAA Atlas NESDIS 84, 35pp.

Garnesson, P., Mangin, A., and Bretagnon M. (2021). Quality Information Document, Ocean Colour Production Centre, Satellite Observation Copernicus-GlobColour Products, https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-OC-QUID-009-030-032-033-037-081-082-083-085-086-098.pdf

Garnier, F., Brankart, J. M., Brasseur, P. and Cosme, E. (2016). Stochastic parameterizations of biogeochemical uncertainties in a 1/4° NEMO/PISCES model for probabilistic comparisons with ocean color data. *Journal of Marine Systems*, 155, 59-72, https://doi.org/10.1016/j.jmarsys.2015.10.012

Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D., Lehodey, P., Perruche, C., Samuelsen, A., and Simon, E. (2015). Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments. *Journal of Operational Oceanography*, 8:sup1, s168-s187, https://doi.org/10.1080/1755876X.2015.1022350

Geider, R.J., MacIntyre, H.L., and Kana, T.M. (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. *Marine Ecology Progress Series*, 148, 187-200, https://doi.org/10.3354/meps148187

Germineaud, C., Brankart, J.M., and Brasseur, P. (2019). An ensemble-based probabilistic score approach to compare observation scenarios: an application to biogeochemical-Argo deployments. *Journal of Atmospheric and Oceanic Technology*, 36(12), 2307-2326, https://doi.org/10.1175/JTECH-D-19-0002.1

Gharamti, M.E., Tjiputra, J., Bethke, I., Samuelsen, A., Skjelvan, I., Bentsen, M., Bertino, L. (2017). Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites. *Ocean Modelling*, 112, 65-89, https://doi.org/10.1016/j.ocemod.2017.02.006

Glibert, P., Mitra, A., Flynn, K., Hansen, P., Jeong, H., and Stoecker, D. (2019). Plants Are Not Animals and Animals Are Not Plants, Right? Wrong! Tiny Creatures in the Ocean Can Be Both at Once! *Frontiers for Young Minds*, 7:48, https://doi.org/10.3389/frym.2019.00048

Gradinger, R. (2009). Sea-ice algae: Major contributors to primary production and algal biomass in the Chukchi and Beaufort Seas during May/June 2002. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56, 1201-1212, https://doi.org/10.1016/j.dsr2.2008.10.016

Green, D.B., Bestley, S., Corney, S.P., Trebilco, R., Lehodey, P., and Hindell, M.A. (2021). Modeling Antarctic krill circumpolar spawning habitat quality to identify regions with potential to support high larval production. *Geophysical Research Letters*, 48, e2020GL091206, https://doi.org/10.1029/2020GL091206

Green, D.B., Bestley, S., Trebilco, R., Corney, S.P., Lehodey, P., McMahon, C.R., Guinet, C., and Hindell, M.A. (2020). Modelled mid-trophic pelagic prey fields improve understanding of marine predator foraging behaviour. *Ecography*, https://doi.org/10.1111/ecog.04939

Gregg, W.W., and Rousseaux, C.S. (2016). Directional and spectral irradiance in ocean models: effects on simulated global phytoplankton, nutrients, and primary production. *Frontiers in Marine Science*, 3, 240, http://dx.doi.org/10.1088/1748-9326/ab4667

Gregg, W.W., and Rousseaux, C.S. (2019). Global ocean primary production trends in the modern ocean color satellite record (1998-2015). *Environmental Research Letters*, 14(12), 124011, https://doi:10.1088/1748-9326/ab4667

Grégoire, M., Raick, C., and Soetaert, K. (2008). Numerical modelling of the central black sea ecosystem functioning during the eutrophication phase. *Progress in Oceanography*, 76, 286-333, https://doi.org/10.1016/j.pocean.2008.01.002

Grégoire, M., Soetaert, K. (2010). Carbon, nitrogen, oxygen and sulfide budgets in the Black Sea: a biogeochemical model of the whole water column coupling the oxic and anoxic parts. *Ecological Modelling*, 221(19), 2287-2301.

Hazen, E. L., Scales, K.L., Maxwell, S.M., Briscoe, D.K., Welch, H., Bograd, S.J., Bailey, H., Benson, S.R., Eguchi, T., Dewar, H., Kohin, S., Costa, D.P., Crowder, L.B., Lewison R.L. (2018). A dynamic ocean management tool to reduce bycatch and support sustainable fisheries.Science Advances, 4(5), DOI: 10.1126/sciadv.aar3001

Hernandez, F., Bertino, L., Brassington, G., Chassignet, E., Cummings, J., Davidson, F., Drevillon, M., Garric, G., Kamachi, M., Lellouche, J.M., Mahdon, R., Martin, M.J,, Ratsimandresy, A., and Regnier, C. (2009). Validation and Inter-comparison studies within GODAE. *Oceanography*, 22(3): 128-143, https://doi.org/10.5670/oceanog.2009.71

Hernandez, F., Smith, G., Baetens, K., Cossarini, G., Garcia-Hermosa, I., Drevillon, M., von Schuckman, K. (2018). Measuring performances, skill and accuracy in operational oceanography: New challenges and approaches. In: "New Frontiers in Operational Oceanography", E. Chassignet, A. Pascual, J. Tintoré, and J. Verron, Eds., GODAE OceanView, 759-796, DOI: 10.17125/gov2018.ch29

Hernandez, O., Lehodey, P., Senina, I., Echevin, V., Ayon, P., Bertrand, A., Gaspar, P. (2014). Understanding mechanisms that control fish spawning and larval recruitment: Parameter optimization of an Eulerian model (SEAPODYM-SP) with Peruvian anchovy and sardine eggs and larvae data. *Progress in Oceanography*, 123, 105-122, http://dx.doi.org/10.1016/j.pocean.2014.03.001

Hipsey, M.R., Gal, G., Arhonditsis, G.B., Carey, C.C., Elliott, J.A., Frassl, M. A., ... & Robson, B.J. (2020). A system of metrics for the assessment and improvement of aquatic ecosystem models. *Environmental Modelling & Software*, 128, 104697, https://doi.org/10.1016/j.envsoft.2020.104697

Hobday, A.J., Hartog, J.R., Timmiss, T., Fielding, J. (2010). Dynamic spatial zoning to manage southern bluefin tuna (Thunnus maccoyii) capture in a multi-species longline fishery. *Fisheries Oceanography*, 19(3), 243-253, https://doi.org/10.1111/j.1365-2419.2010.00540.x

Houtekamer, P. L., and Mitchell, H. L. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, 126(3), 796-811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2

Howell, E.A., Kobayashi, D.R., Parker, D.M., Balazs, G.H., Polovina, J.J. (2008). TurtleWatch: a tool to aid in the bycatch reduction of loggerhead turtles Caretta caretta in the Hawaii-based pelagic longline fishery. *Endangered Species Research*, 5:267-278, https://doi.org/10.3354/esr00096

Hu, J., Fennel, K., Mattern, J.P. and Wilkin, J. (2012). Data assimilation with a local Ensemble Kalman Filter applied to a three-dimensional biological model of the Middle Atlantic Bight. *Journal of Marine Systems*, 94, 145-156, https://doi.org/10.1016/j.jmarsys.2011.11.016

Huse, G., Melle, W., Skogen, M.D., Hjøllo, S.S., Svendsen, E., and Budgell, W.P. (2018). Modeling Emergent Life Histories of Copepods. Frontiers in Ecology and Evolution, 6(23), https://doi.org/10.3389/fevo.2018.00023

IOCCG. (2020). Synergy between Ocean Colour and Biogeochemical/Ecosystem Models. Dutkiewicz, S. (ed.), IOCCG Report Series, No. 19, International Ocean Colour Coordinating Group, Dartmouth, Canada, http://dx.doi.org/10.25607/OBP-711

Ishizaka, J. (1990). Coupling of coastal zone color scanner data to a physical-biological model of the southeastern US continental shelf ecosystem: 2. An Eulerian model. *Journal of Geophysical Research: Oceans*, 95(C11), 20183-20199, https://doi.org/10.1029/JC095iC11p20183

Jaccard, P., Hjermann, D. Ø., Ruohola, J., Marty, S., Kristiansen, T., Sørensen, K., Kaitala, S., Mangin, A., Pouliquen, S., et al. (2021). Quality Information Document For Global Ocean Reprocessed in-situ Observations of Biogeochemical Products INSITU_GLO_BGC_REP_OBSERVATIONS_013_046, http://doi.org/10.13155/54846, https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-INS-QUID-013-046.pdf

Jech, J.M., Horne, J.K., Chu, D., Demer, D.A., Francis, D.T.I., Gorska, N., Jones, B., Lavery, A.C., Stanton, T.K., Macaulay, G.J., Reeder, D.B., Sawada, K. (2015). Comparisons among ten models of acoustic backscattering used in aquatic ecosystems. *The Journal of the Acoustical Society of America*, 138: 3742, https://doi.org/10.1121/1.4937607

Jennings, S., and Collingridge, K. (2015). Predicting Consumer Biomass, Size-Structure, Production, Catch Potential, Responses to Fishing and Associated Uncertainties in the World's Marine Ecosystems. *PLoS ONE*, 10(7), e0133794. https://doi.org/10.1371/journal.pone.0133794

Jolliff, J.K., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M.A.M., Helber, R.W., and Arnone, R. (2009). Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems*, 76, 64-82, https://doi.org/10.1016/j.jmarsys.2008.05.014

Jones, E.M., Baird, M.E., Mongin, M., Parslow, J., Skerratt, J., Lovell, J., Margvelashvili, N., Matear, R.J., Wild-Allen, K., Robson, B. and Rizwi, F. (2016). Use of remote-sensing reflectance to constrain a data assimilating marine biogeochemical model of the Great Barrier Reef. *Biogeosciences*, 13(23), 6441, https://doi.org/10.5194/bg-13-6441-2016

Kaufman, D.E. (2017). Using High-Resolution Glider Data and Biogeochemical Modeling to Investigate Phytoplankton Variability in the Ross Sea. Dissertations, Theses, and Masters Projects, William & Mary, Paper 1499449869, http://dx.doi.org/10.21220/M2BK8V

Kriest, I., Kähler, P., Koeve, W., Kvale, K., Sauerland, V., and Oschlies, A. (2020). One size fits all? Calibrating an ocean biogeochemistry model for different circulations. *Biogeosciences*, 17, 3057-3082, https://doi.org/10.5194/bg-17-3057-2020

Krishnamurthy, A., Moore, J. K., Mahowald, N., Luo, C., and Zender, C.S. (2010). Impacts of atmospheric nutrient inputs on marine biogeochemistry. *Journal of Geophysical Research: Biogeosciences*, 115(G1), https://doi.org/10.1029/2009JG001115

Lambert, C., Mannocci, L., Lehodey, P., Ridoux, V. (2014). Predicting Cetacean Habitats from Their Energetic Needs and the distribution of Their Prey in Two Contrasted Tropical Regions. PLoS ONE, 9(8), e105958, https://doi.org/10.1371/journal.pone.0105958

Lamouroux, J., Perruche, C., Mignot, A., Paul, J., Szczypta, C. (2019). Quality Information Document For Global Biogeochemical Analysis and Forecast Product, https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-GLO-QUID-001-028.pdf

Lamouroux, J., Perruche, C., Mignot, A., Gutknecht, E., Ruggiero, G., Evaluation of the CMEMS global biogeochemical simulation, with assimilation of satellite Chla concentrations, in prep.

Laws, E.A. (2013). Evaluation of In Situ Phytoplankton Growth Rates: A Synthesis of Data from Varied Approaches. *Annual Review of Marine Science*, 5(1), 247-268, https://doi.org/10.1146/annurev-marine-121211-172258

Le Quéré, C., Harrison, S. P., Prentice, I. C., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Cotrim Da Cunha, L., Geider, R., Giraud, X., Klaas, C., Kohfeld, K. E., Legendre, L., Manizza, M., Platt, T., Rivkin, R. B., Sathyendranath, S., Uitz, J., Watson, A. J., and Wolf?Gladrow, D. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, 11, 2016-2040, https://doi.org/10.1111/j.1365-2486.2005.1004.x

Le Traon et al. (2017). The Copernicus marine environmental monitoring service: main scientific achievements and future prospects. Special Issue Mercator Océan International #56. Available at: https://marine.copernicus.eu/it/node/594

Le Traon, P. Y., Reppucci, A., Fanjul, E. A., Aouf, L., Behrens, A., Belmonte, M., Bentamy, A., Bertino, L., Brando, V. E., Kreiner, M. B., Benkiran, M., Carval, T., Ciliberti, S. A., Claustre, H., Clementi, E., Coppini, G., Cossarini, G., De Alfonso Alonso- Muñoyerro, M., Delamarche, A., Dibarboure, G., Dinessen, F., Drevillon, M., Drillet, Y., Faugere, Y., Fernández, V., Fleming, A., Garcia-Hermosa, M. I., Sotillo, M. G., Garric, G., Gasparin, F., Giordan, C., Gehlen, M., Grégoire, M., Guinehut, S., Hamon, M., Harris, C., Hernandez, F., Hinkler, J. B., Hoyer, J., Karvonen, J., Kay, S., King, R., Lavergne, T., Lemieux-Dudon, B., Lima, L., Mao, C., Martin, M. J., Masina, S., Melet, A., Nardelli, B. B., Nolan, G., Pascual, A., Pistoia, J., Palazov, A., Piolle, J. F., Pujol, M. I., Pequignet, A. C., Peneva, E., Gómez, B. P., de la Villeon, L. P., Pinardi, N., Pisano, A., Pouliquen, S., Reid, R., Remy, E., Santoleri, R., Siddorn, J., She, J., Staneva, J., Stoffelen, A., Tonani, M., Vandenbulcke, L., von Schuckmann, K., Volpe, G., Wettre, C., and Zacharioudaki, A. (2019). From observation to information and users: The Copernicus Marine Service Perspective. *Frontiers in Marine Science*, https://doi.org/10.3389/fmars.2019.00234

Lehodey, P., Senina, I., Murtugudde, R. (2008). A Spatial Ecosystem And Populations Dynamics Model (SEAPODYM) - Modelling of tuna and tuna-like populations. *Progress in Oceanography*, 78, 304-318, https://doi.org/10.1016/j.pocean.2008.06.004

Lehodey, P., Senina, I., Wibawa, T.A., Titaud, O., Calmettes, B., Tranchant, B., and Gaspar, P. (2017). Operational modelling of bigeye tuna (Thunnus obesus) spatial dynamics in the Indonesian region. *Marine Pollution Bulletin*, 131, 19-32, https://doi.org/10.1016/j.marpolbul.2017.08.020

Lehodey, P., Conchon, A., Senina, I., Domokos, R., Calmettes, B., Jouanno, J., Hernandez, O., and Kloser, R. (2015). Optimization of a micronekton model with acoustic data. I*CES Journal of Marine Science*, 72(5), 1399-1412, https://doi.org/10.1093/icesjms/fsu233

Lehodey, P., Murtugudde, R., and Senina, I. (2010). Bridging the gap from ocean models to population dynamics of large marine predators: A model of mid-trophic functional groups. Progress in Oceanography, 84, 69-84.

Lehodey, P., Senina, I., Calmettes, B, Hampton, J, Nicol S. (2013). Modelling the impact of climate change on Pacific skipjack tuna population and fisheries. Climatic Change, 119 (1): 95-109, DOI 10.1007/s10584-012-0595-1

Lengaigne, M., Menkes, C., Aumont, O., Gorgues, T., Bopp, L., André, J. M., and Madec, G. (2007). Influence of the oceanic biology on the tropical Pacific climate in a coupled general circulation model. *Climate Dynamics*, 28(5), 503-516. https://doi.org/10.1007/s00382-006-0200-2

Libralato, S., and Solidoro, C. (2009). Bridging biogeochemical and food web models for an End-to-End representation of marine ecosystem dynamics: The Venice lagoon case study. *Ecological Modelling*, 220(21), 2960-2971, https://doi.org/10.1016/j.ecolmodel.2009.08.017

Longhurst, A. (1998). Ecological geography in the sea. Academic Press.

Mattern, J. P., Fennel, K., and Dowd, M. (2012). Estimating time-dependent parameters for a biological ocean model using an emulator approach. *Journal of Marine Systems*, 96, 32-47, https://doi.org/10.1016/j.jmarsys.2012.01.015

Mattern, J.P., Dowd, M. and Fennel, K. (2013). Particle filter-based data assimilation for a three-dimensional biological ocean model and satellite observations. *Journal of Geophysical Research: Oceans*, 118(5), 2746-2760, https://doi.org/10.1002/jgrc.20213

Maunder, M.N., Punt, A.E. (2013). A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, 142, 61-74, https://doi.org/10.1016/j.fishres.2012.07.025

Maury, O., Faugeras, B., Shin, Y.-J., Poggiale, J.-C., Ben Ari, T., and Marsac, F. (2007). Modeling environmental effects on the size-structured energy flow through marine ecosystems. Part 1: The model. *Progress in Oceanography*, 74(4), 479-499, https://doi.org/10.1016/j.pocean.2007.05.002

McEwan, R., S. Kay, D. Ford (2021). Quality Information Document For the Atlantic - European North West Shelf Production Centre, https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-NWS-QUID-004-002.pdf, https://doi.org/10.48670/moi-00056

Melsom, A. and Ç. Yumruktepe (2021), Quality Information Document For the Arctic Production Centre, https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-ARC-QUID-002-004.pdf, https://doi.org/10.48670/moi-00003

Mignot, A., Claustre, H., Cossarini, G., D'Ortenzio, F., Gutknecht, E., Lamouroux, J., Lazzari, P., Perruche, C., Salon, S., Sauzède, R., Taillandier, V., Teruzzi, A. (2021). Defining BGC-Argo-based metrics of ocean health and biogeochemical functioning for the evaluation of global ocean models. *Biogeosciences*, https://doi.org/10.5194/bg-2021-2

Miller, C.B., Lynch, D.R., Carlotti, F., Gentleman, W., Lewis, C.V.W. (1998). Coupling of an individual-based population dynamic model of Calanus finmarchicus to a circulation model for the Georges Bank region. *Fisheries Oceanography*, 7(3-4), 219-234, https://doi.org/10.1046/j.1365-2419.1998.00072.x

Mogensen, K.S., Balmaseda, M.A., Weaver, A., Martin, M.J., Vidard, A. (2009). NEMOVAR: A variational data assimilation system for the NEMO ocean model. ECMWF Newsl., 120, 17-21.

Moore, A. M., Martin, M. J., Akella, S., Arango, H. G., Balmaseda, M., Bertino, L., ... and Weaver, A. T. (2019). Synthesis of ocean observations using data assimilation for operational, real-time and reanalysis systems: A more complete picture of the state of the ocean. Frontiers in Marine Science, 6:90, https://doi.org/10.3389/fmars.2019.00090

Moriarty, R., and O'Brien, T. (2013). Distribution of mesozooplankton biomass in the global ocean. *Earth System Science Data*, 5(1), 45-55, https://doi.org/10.5194/essd-5-45-2013

Natvik, L., and Evensen, G. (2003). Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 1: Data assimilation experiments. *Journal of Marine Systems*, 41, 127-153, https://doi.org/10.1016/S0924-7963(03)00016-2

Nerger, L. and Gregg, W.W. (2007). Assimilation of SeaWiFS data into a global ocean-biogeochemical model using a local SEIK filter. *Journal of Marine Systems*, 68, 237-254, https://doi.org/10.1016/j.jmarsys.2006.11.009

Nerger, L., and Gregg, W.W. (2008). Improving assimilation of SeaWiFS data by the application of bias correction with a local SEIK filter. *Journal of Marine Systems*, 73, 87-102, https://doi.org/10.1016/j.jmarsys.2007.09.007

Nerger, L., Janjić, T., Schröter, J., and Hiller, W. (2012). A Unification of Ensemble Square Root Kalman Filters. *Monthly Weather Review*, 140(7), 2335-2345, https://doi.org/10.1175/MWR-D-11-00102.1

Neumann, T. (2000). Towards a 3D-ecosystem model of the Baltic Sea. *Journal of Marine Systems*, 25(3), 405-419, https://doi.org/10.1016/S0924-7963(00)00030-0

Olsen, A., Lange, N., Key, R. M., Tanhua, T., Bittig, H. C., Kozyr, A., Àlvarez, M., Azetsu-Scott, K., Becker, S., Brown, P. J., Carter, B. R., Cotrim da Cunha, L., Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Jeansson, E., Jutterström, S., Landa, C. S., Lauvset, S. K., Michaelis, P., Murata, A., Pérez, F. F., Pfeil, B., Schirnick, C., Steinfeldt, R., Suzuki, T., Tilbrook, B., Velo, A., Wanninkhof, R. and Woosley, R. J. (2020). GLODAPv2.2020 - the second update of GLODAPv2. *Earth System Science Data*, https://doi.org/10.5194/essd-12-3653-2020

Ourmières, Y., Brasseur, P., Lévy, M., Brankart, J.M. and Verron, J. (2009). On the key role of nutrient data to constrain a coupled physical-biogeochemical assimilative model of the North Atlantic Ocean. *Journal of Marine Systems*, 75(1-2), 100-115, https://doi.org/10.1016/j.jmarsys.2008.08.003

Palmer, J. R., and Totterdell, I. J. (2001). Production and export in a global ocean ecosystem model. *Deep Sea Research Part I: Oceanographic Research Papers*, 48(5), 1169-1198, https://doi.org/10.1016/S0967-0637(00)00080-7

Park, J.-Y., Stock, C. A., Yang, X., Dunne, J. P., Rosati, A., John, J., et al. (2018). Modeling global ocean biogeochemistry with physical data assimilation: A pragmatic solution to the equatorial instability. *Journal of Advances in Modeling Earth Systems*, 10, 891- 906, https://doi.org/10.1002/2017MS001223

Pérez-Jorge, S., Tobeña, M., Prieto, R., Vandeperre, F., Calmettes, B., Lehodey, P., Silva, M.A. (2020). Environmental drivers of large-scale movements of baleen whales in the mid-North Atlantic Ocean. *Diversity and Distributions*, 26(6), 683-698, https://doi.org/10.1111/ddi.13038

Petrik, C.M., Stock, C.A., Andersen, K.H., van Denderen, P.D., and Watson, J.R. (2019). Bottom-up drivers of global patterns of demersal, forage, and pelagic fishes. *Progress in Oceanography*, 176:102124, https://doi.org/10.1016/j.pocean.2019.102124

Petrik, C.M., Stock, C.A., Andersen, K.H., van Denderen, P.D., and Watson, J.R. (2020). Large pelagic fish are most sensitive to climate change despite pelagification of ocean foodwebs. *Frontiers in Marine Science*, 7, 588482, https://doi.org/10.3389/fmars.2020.588482

Pham, D.T., Verron, J., Roubaud, M.C. (1998). A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16(3-4): 323-340, https://doi.org/10.1016/S0924-7963(97)00109-7

Pradhan, H.K., Völker, C., Losa, S.N., Bracher, A., and Nerger, L. (2020). Global assimilation of ocean-color data of phytoplankton functional types: Impact of different data sets. *Journal of Geophysical Research: Oceans*, 125, e2019JC015586, https://doi.org/10.1029/2019JC015586

Proud, R., Handegard, N. O., Kloser, R. J., Cox, M. J., and Brierley, A. S. (2018). From siphonophores to deep scattering layers:Uncertainty ranges for the estimation of global mesopelagic fish biomass. *ICES Journal of Marine Science*, 76, 718-733, https://doi.org/10.1093/icesjms/fsy037

Redfield, A.C. (1934). On the proportions of organic derivatives in sea water and their relation to the composition of plankton. James Johnson Memorial Volume, R. J. Daniel, Ed., University Press of Liverpool, 177-192.

Roberts, J. J., Best, B. D., Mannocci, L., Fujioka, E., Halpin, P. N., Palka, D. L., ... Lockhart, G. G. (2016). Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. *Nature Scientific Reports*, 6, 22615, https://doi.org/10.1038/srep22615

Romagosa M., Lucas C., Pérez-Jorge S., Tobeña M., Lehodey P., Reis J., Cascão I., Lammers M. O., Caldeira R. M. A., Silva M. A. (2020). Differences in regional oceanography and prey biomass influence the presence of foraging odontocetes at two Atlantic seamounts. *Marine Mammal Science*, 36: 158-179, https://doi.org/10.1111/mms.12626

Romagosa, M., Pérez-Jorge, S., Cascão, I., Mouriño, H., Lehodey, P., Marques, T.A, Silva, M.A. (2021). Food talk: 40-Hz fin whale calls are associated with prey availability. *Proceedings of the Royal Society B Biological Sciences*, 288(1954), https://doi.org/10.1098/rspb.2021.1156

Rose, K.A., Fiechter, J., Curchitser, E.N., Hedstrom, K., Bernal, M., Creekmore, S., Haynie, A.C., Ito, S., Lluch-Cota, S.E., Megrey, B.A., Edwards, C.A., Checkley, D.M., Koslow, T., Mcclatchie, S., Werner, F.E., Maccall, A.D., and Agostini, V.N. (2015). Demonstration of a fully-coupled end-to-end model for small pelagic fish using sardine and anchovy in the California Current. *Progress in Oceanography*, 138, 348-380, https://doi.org/10.1016/j.pocean.2015.01.012

Russell, J., Sarmiento, J., Cullen, H., Hotinski, R., Johnson, K.S., Riser, S.C., et al. (2014). The Southern Ocean Carbon and Climate Observations and Modeling Program (SOCCOM). Ocean Carbon Biogeochem. News. Available at https://web.whoi.edu/ocb/wp-content/uploads/sites/43/2016/12/OCB_NEWS_FALL14.pdf

Ryabinin, V., Barbière, J., Haugan, P., Kullenberg, G., Smith, N., McLean, C., Troisi, A., Fischer, A., Aricò, S., Aarup, T., Pissierssens, P., Visbeck, M., Enevoldsen, H.O., Rigaud, J. (2019). The UN Decade of Ocean Science for Sustainable Development, *Frontiers in Marine Science*, 6,470, https://doi.org/10.3389/fmars.2019.00470

Sakov, P., and Oke, P.R. (2008). A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus* A, 60, 361-371, https://doi.org/10.1111/j.1600-0870.2007.00299.x

Salon, S., Cossarini, G., Bolzon, G., Feudale, L., Lazzari, P., Teruzzi, A., Solidoro, C., Crise, A. (2019). Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts. *Ocean Science*, 15, 997-1022, https://doi.org/10.5194/os-15-997-2019

Santana-Falcon, Y., Brasseur, P., Brankart, J.M., and Garnier, F. (2020). Assimilation of chlorophyll data into a stochastic ensemble simulation for the North Atlantic ocean, *Ocean Science*, 16, 1297-1315, https://doi.org/10.5194/os-16-1297-2020

Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A. (2017). Reviews and syntheses: parameter identification in marine planktonic ecosystem modelling. *Biogeosciences*, 14, 1647-1701, https://doi.org/10.5194/bg-14-1647-2017

Scutt Phillips, J., Sen Gupta, A., Senina, I., van Sebille, E., Lange, M., Lehodey, P., Hampton, J., Nicol, S. (2018). An individual-based model of skipjack tuna (Katsuwonus pelamis) movement in the tropical Pacific Ocean. *Progress in Oceanography*, 164, 63-74, https://doi.org/10.1016/j.pocean.2018.04.007

Senina, I., Lehodey, P., Hampton, J., Sibert, J. (2019). Quantitative modelling of the spatial dynamics of South Pacific and Atlantic albacore tuna populations. *Deep Sea Research Part II: Topical Studies in Oceanography*, 175, 104667, https://doi.org/10.1016/j.dsr2.2019.104667

Senina, I., Lehodey, P., Sibert, J., Hampton, J. (2020). Improving predictions of a spatially explicit fish population dynamics model using tagging data. *Canadian Journal of Aquatic and Fisheries Sciences*, 77(3), 576-593, https://doi.org/10.1139/cjfas-2018-0470

Senina, I., Sibert, J., and Lehodey, P. (2008). Parameter estimation for basin-scale ecosystem-linked population models of large pelagic predators: Application to skipjack tuna. *Progress in Oceanography*, 78, 319-335, https://doi.org/10.1016/j.pocean.2008.06.003

Sibert, J., Senina, I., Lehodey, P., Hampton, J. (2012). Shifting from marine reserves to maritime zoning for conservation of Pacific bigeye tuna (Thunnus obesus). *Proceedings of the National Academy of Sciences*, 109(44): 18221-18225.

Simon, E. and Bertino, L. (2009). Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. *Ocean Science*, 5(4), 495-510, https://doi.org/10.5194/os-5-495-2009

Simon, E. Bertino, L. (2012). Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model. *Journal of Marine Systems*, 89(1), 1-18, https://doi.org/10.1016/j.jmarsys.2011.07.007

Simon, E., Samuelsen, A., Bertino, L. and Mouysset, S. (2015). Experiences in multiyear combined state-parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using the Ensemble Kalman Filter. *Journal of Marine Systems*, 152, 1–17, https://doi.org/10.1016/j.jmarsys.2015.07.004

Skákala, J., Ford, D., Brewin, R. J. W., McEwan, R., Kay, S., Taylor, B., Mora, L., and Ciavatta, S. (2018). The assimilation of phytoplankton functional types for operational forecasting in the northwest European shelf. *Journal of Geophysical Research: Oceans*, 123, 5230-5247, https://doi.org/10.1029/2018JC014153

Skakala, J., Bruggeman, J., Brewin, R.J., Ford, D.A. and Ciavatta, S. (2020). Improved representation of underwater light field and its impact on ecosystem dynamics: A study in the North Sea. *Journal of Geophysical Research: Oceans*, 125(7), e2020JC016122, https://doi.org/10.1029/2020JC016122

Skákala, J., Ford, D., Bruggeman, J., Hull, T., Kaiser, J., King, R.R., Loveday, B., Palmer, M.R., Smyth, T., Williams, C.A. and Ciavatta, S. (2021a). Towards a multi-platform assimilative system for North Sea biogeochemistry. *Journal of Geophysical Research: Oceans*, 126(4), e2020JC016649, https://doi.org/10.1029/2020JC016649

Skakala, J., Bruggeman, J., Ford, D.A., Wakelin, S.L., Akpınar, A., Hull, T., Kaiser, J., Loveday, B.R., Williams, C.A.J. and Ciavatta, S. (2021b). Improved consistency between the modelling of ocean optics, biogeochemistry and physics, and its impact on the North-West European Shelf seas. *Earth and Space Science Open Archive ESSOAr*, https://doi.org/10.1002/essoar.10506737.2

Skogen, M.D. (1993). A User's guide to NORWECOM (the NORWegian ECOlogical Model system). Technical report 6, Inst.of Marine Research, Division of Marine Env., Pb 1870, N-5024 Bergen, Norway.

Skogen, M.D. and Søiland, H. (1998). A User's guide to NORWECOM v2.0 (the NORWegian ECOlogical Model system). Tech.rep. Fisken og Havet 18, Inst. of Marine Research, Pb.1870, N-5024 Bergen. Norway.

Song, H., Edwards, C.A., Moore, A.M., and Fiechter, J. (2016). Data assimilation in a coupled physical-biogeochemical model of the California current system using an incremental lognormal 4-dimensional variational approach: part 3 - Assimilation in a realistic context using satellite and in situ observations. Ocean Modelling, 106, 1531-145, https://doi.org/10.1016/j.ocemod.2016.04.001

Spindler, M. (1994). Notes on the biology of sea ice in the Arctic and Antarctic. *Polar Biology*, 14, 319-324, https://doi.org/10.1007/BF00238447

Storto, A., Oddo, P., Cipollone, A., Mirouze, I. and Lemieux-Dudon, B. (2018). Extending an oceanographic variational scheme to allow for affordable hybrid and four-dimensional data assimilation. *Ocean Modelling*, 128,.67-86, https://doi.org/10.1016/j.ocemod.2018.06.005

Stow, C. A., Jolliff, J., McGillicuddy Jr, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems*, 76(1-2), 4-15, https://doi.org/10.1016/j.jmarsys.2008.03.011

Taylor, K.E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183-7192, https://doi.org/10.1029/2000JD900719

Teruzzi, A, Di Cerbo, P, Cossarini, G, Pascolo, E, Salon, S. (2019). Parallel implementation of a data assimilation scheme for operational oceanography: The case of the MedBFM model system. *Computers & Geosciences*, 124, 103-114, https://doi.org/10.1016/j.cageo.2019.01.003

Teruzzi, A., Dobricic, S., Solidoro, C., and Cossarini, G. (2014). A 3-D variational assimilation scheme in coupled transport-biogeochemical models: Forecast of Mediterranean biogeochemical properties. *Journal of Geophysical Research: Oceans*, 119, 200-217, https://doi.org/10.1002/2013JC009277

Tissier, A.-S., Brankart, J.M., Testut, C.E., Ruggiero, G., Cosme, E., and Brasseur, P. (2019). A multiscale ocean data assimilation approach combining spatial and spectral localization. *Ocean Science*, 15, 443-457, https://doi.org/10.5194/os-15-443-2019

Torres, R., Allen, J.I. and Figueiras, F.G. (2006). Sequential data assimilation in an upwelling influenced estuary. *Journal of Marine Systems*, 60(3-4), 317-329, https://doi.org/10.1016/j.jmarsys.2006.02.001

Uitz, J., Claustre, H., Morel, A., and Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research: Oceans*, 111(C8), https://doi.org/10.1029/2005JC003207

van Leeuwen, P.J. (2010). Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653), 1991-1999, https://doi.org/10.1002/qj.699

Vandenbulcke, L., and Barth, A. (2015). A stochastic operational system of the Black Sea: Tech. and validation. *Ocean Modelling*, 93, 7-21.

Verdy, A., and Mazloff, M.R. (2017). A data assimilating model for estimating Southern Ocean biogeochemistry. *Journal of Geophysical Research: Oceans*, 122(9), 6968-6988, https://doi.org/10.1016/j.ocemod.2015.07.010

Vetra-Carvalho, S., Van Leeuwen, P.J., Nerger, L., Barth, A., Altaf, M.U., Brasseur, P., Kirchgessner, P. and Beckers, J.M. (2018). State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus A: Dynamic Meteorology and Oceanography*, 70(1), 1-43, https://doi.org/10.1080/16000870.2018.1445364

Vichi, M., Cossarini, G., Gutierrez Mlot, E., Lazzari, P., Lovato, T., Mattia, G., Masina, S., McKiver, W., Pinardi, N., Solidoro, C., Zavatarelli, M. (2015). The Biogeochemical Flux Model (BFM): Equation Description and User Manual. BFM version 5.1. BFM Report series N.1, March 2015, Bologna, Italy, pp. 89.

Wang, B., Fennel, K., Yu, L. and Gordon, C. (2020). Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico. *Biogeosciences*, 17(15), 4059-4074, https://doi.org/10.5194/bg-17-4059-2020

Warner, J.C., Armstrong, B., He, R., Zambon, J.B. (2010). Development of a Coupled Ocean–Atmosphere–Wave–Sediment Transport (COAWST) Modeling System. *Ocean Modelling*, 35(3), 230-244, https://doi.org/10.1016/j.ocemod.2010.07.010

Waters, J., Lea, D.J., Martin, M.J., Mirouze, I., Weaver, A. and While, J. (2015). Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. *Quarterly Journal of the Royal Meteorological Society*, 141(687), 333-349, https://doi.org/10.1002/qj.2388

Waters, J., Bell, M. J., Martin, M. J., and Lea, D. J. (2017). Reducing ocean model imbalances in the equatorial region caused by data assimilation. *Quarterly Journal of the Royal Meteorological Society, 143(702), 195-208.*

While, J., Totterdell, I. and Martin, M. (2012). Assimilation of pCO2 data into a global coupled physical-biogeochemical ocean model. *Journal of Geophysical Research: Oceans*, 117(C3), https://doi.org/10.1029/2010JC006815

Wright, R.M., Le Quéré, C., Buitenhui, E., Pitois, S., Gibbons, M. (2021). Role of jellyfish in the plankton ecosystem revealed using a global ocean biogeochemical model. *Biogeosciences*, 18, 1291-1320, https://doi.org/10.5194/bg-18-1291-2021

Xiao, Y., and M.A.M. Friedrichs (2014). Using biogeochemical data assimilation to assess the relative skill of multiple ecosystem models: effects of increasing the complexity of the planktonic food web. *Biogeosciences*, 11(11), 3015-3030, https://doi.org/10.5194/bg-11-3015-2014

Yool, A., Popova, E. E., and Anderson, T. R. (2013). MEDUSA-2.0: an intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies. *Geoscientific Model Development*, 6, 1767-1811, https://doi.org/10.5194/gmd-6-1767-2013

Yu, L., Fennel, K., Bertino, L., El Gharamti, M., and Thompson, K. R. (2018). Insights on multivariate updates of physical and biogeochemical ocean variables using an Ensemble Kalman Filter and an idealized model of upwelling. *Ocean Modelling*, 126, 13-28, doi:10.1016/j.ocemod.2018.04.005

Zeebe, R., and Wolf-Gladrow, D. (2001). CO2 in Seawater: Equilibrium, Kinetics, Isotopes. Elsevier Oceanography Book Series, 65, 346 pp, Amsterdam.

Zhou M., Carlotti F., Zhu Y. (2010). A size-spectrum zooplankton closure model for ecosystem modelling. *Journal of Plankton Research*, 32(8), 1147-1165, https://doi.org/10.1093/plankt/fbq054