

An Ensemble-Based Probabilistic Score Approach to Compare Observation Scenarios: An Application to Biogeochemical-Argo Deployments

CYRIL GERMINEAUD,^a JEAN-MICHEL BRANKART, AND PIERRE BRASSEUR

Université Grenoble Alpes, CNRS, IRD, Grenoble-INP, IGE, Grenoble, France

(Manuscript received 7 January 2019, in final form 5 August 2019)

ABSTRACT

A cross-validation algorithm is developed to perform probabilistic observing system simulation experiments (OSSEs). The use of a probability distribution of “true” states is considered rather than a single “truth” using a cross-validation algorithm in which each member of an ensemble simulation is alternatively used as the “truth” and to simulate synthetic observation data that reflect the observing system to be evaluated. The other available members are used to produce an updated ensemble by assimilating the specific data, while a probabilistic evaluation of the observation impacts is obtained using a comprehensive set of verification skill scores. To showcase this new type of OSSE studies with tractable numerical costs, a simple biogeochemical application under the Horizon 2020 AtlantOS project is presented for a single assimilation time step, in order to investigate the value of adding biogeochemical (BGC)-Argo floats to the existing satellite ocean color observations. Further experiments must be performed in time as well for a rigorous and effective evaluation of the BGC-Argo network design, though some evidence from this preliminary work suggests that assimilating chlorophyll data from a BGC-Argo array of 1000 floats can provide additional error reduction at the surface, where the use of spatial ocean color data is limited (due to cloudy conditions), as well at depths ranging from 50 to 150 m.


1. Introduction

The global ocean observing system is based on various in situ and satellite components that are mostly intermittent and loosely connected as they often result from monodisciplinary initiatives led by national and/or international agencies. This lack of integration between the observing components was outlined during the OceanObs’09 (see www.oceanobs09.net) conference, along with the societal needs for a sustained ocean observing system. Accordingly, a task team was formed to develop a framework that can guide the future implementation of a better-coordinated and sustained

global observing system related to both climate and marine ecosystems.

Following the path traced by OceanObs’09, a 5-yr international collaboration was created in 2014, the Horizon 2020 AtlantOS project ([Visbeck et al. 2015](#); see the [appendix](#) for expansions of some acronyms used in the text). This collaboration aims to enhance and redesign the existing observing system in the Atlantic Ocean, and to produce multidisciplinary and sustainable datasets that will be collectively shared and will satisfy the needs of various end-users. As part of AtlantOS, dedicated efforts have been conducted to investigate integrated observing systems that involve satellite, in situ observations from new technologies (or networks) and data-assimilating models, as previously tested in international programs such as GODAE OceanView (see [Bell et al. 2015](#)).

In meteorology and more recently in oceanography, the most common method used to evaluate the impact of observations, is to perform data-denial experiments using a data-assimilative simulation run. These experiments, known as observing system experiments (OSEs), aim to assess what happens when specific observation data are removed from or added to the data-assimilative

 Denotes content that is immediately available upon publication as open access.

^a Current affiliation: Cooperative Institute for Marine and Atmospheric Studies, University of Miami, and NOAA/Atlantic Oceanographic and Meteorological Laboratory/Physical Oceanography Division, Miami, Florida.

Corresponding author: Cyril Germineaud, cyril.germineaud@noaa.gov

run. To evaluate the impact of these removed/added data, comparisons are made with a reference run in which the tested data are not assimilated (e.g., [Fujii et al. 2015a,b](#); [Oke et al. 2015](#); [Xue et al. 2017](#)).

When OSEs are applied to new types of data or nonexistent observing systems, they are referred to as observing system simulation experiments (OSSEs). The observation data to be assimilated are synthetic with this approach; they are simulated from a non-assimilative run that is assumed to represent the “true” state of the system, known as the “nature run.” The impact of the synthetic data is then assessed against this “true state” for each conducted experiment. OSSEs are typically used to examine the performance of future observing systems (e.g., [Alvarez and Mourre 2014](#); [Atlas et al. 2015](#)), help for deciding between different competing instrument configurations (e.g., [Hoffman and Atlas 2016](#)), and have proved to be a cost-effective approach to compare different deployment strategies (e.g., [Halliwell et al. 2015](#)). Another motivation for OSSEs is to test the impact of assimilating new observation types on weather (or oceanic) forecasts, and provide opportunities for improving operational data assimilation systems used in numerical weather prediction (NWP) centers (e.g., [Halliwell et al. 2017](#)).

Within AtlantOS, objective recommendations to enhance the Atlantic observing system and implement new components for ocean physics and marine biogeochemistry are given using the OSSE approach. This activity relies on various modeling and assimilation systems developed by the European research community ([Gasparin et al. 2019](#)). The optimal observing strategy for the different components is examined using coordinated scenarios among the various AtlantOS groups who have been conducting the OSSEs. For the physical variables, these model-based studies involve a single simulated “true” ocean, which is assumed to realistically represent the ocean physics variability (in both space and time). The various tested observing system designs are based on observations simulated from the nature run. Different data types are generated using Argo float profiles and drifting buoy trajectories, moorings at fixed locations, and satellite data coverage of sea surface height and sea surface temperature.

For biogeochemical variables, the in situ observing system is still underdeveloped compared to ocean physics, delivering only scattered and uneven data coverage based on sparse ship cruises, glider experiments, and fixed moorings. So far, sufficient data coverage has only been realized by ocean color satellite missions, which have helped to better understand and monitor observed phenomena such as primary production

variability and bloom formation. The only widespread source of valuable observations for assimilating biogeochemical data is thus limited to the sea surface chlorophyll concentrations, as ocean color sensors do not measure other biogeochemical variables (e.g., nutrients or trophic species).

In close synergy with ocean color satellites, a global array of biogeochemical sensors analogous to the existing core Argo network would revolutionize our knowledge of the changing state of primary productivity, ocean carbon cycling, acidification and the patterns of marine ecosystems variability from seasonal to inter-annual time scales. To implement this biogeochemical Argo array, several pilot experiments [see [Johnson and Claustre \(2016b\)](#) for more detail] were performed to test prototype profiling float arrays equipped with various biogeochemical sensors (e.g., chlorophyll, nitrate, pH, oxygen). Those experiments have shown the observational richness of having a biogeochemical-Argo (BGC-Argo) network at regional scales, though the deployment strategy of such an array at global scale remains under investigation, including its interaction with other components of the observing system.

The future BGC-Argo data (with sufficient coverage) together with ocean color observations will be assimilated into coupled physical–biogeochemical models, allowing a new generation of biogeochemical forecasting systems in tight connection with the EU Copernicus Marine Environment Monitoring Service (CMEMS), as well as major developments in data-assimilation and modeling experiments (as achieved with Argo’s physical observations under the GODAE OceanView initiative). Since the last decade, coupled physical–biogeochemical models associated with ocean color data assimilation schemes have progressed toward the objective of providing faithful monitoring and prediction of the near-surface biogeochemistry state of the ocean (e.g., [Ford and Barciela 2017](#); [Ciavatta et al. 2018](#); [Skákala et al. 2018](#)). As BGC-Argo floats will extend biogeochemical observations to the subsurface ocean, further improvements are expected to be made in the next 10 years.

Nevertheless, the lack of clear “principles” governing the evolution of marine ecosystems and inevitable computational limitations require many simplifications in the biogeochemical models equations. These simplifications lead to poor (or even omitted) representations of processes, and therefore to significant model uncertainties that limit the predictability and monitoring of the system. One element of progress is the transition to ensemble simulations to simulate model uncertainties that can be, for example, associated with unresolved biological diversity (e.g., [Brankart et al. 2015](#))

or key biogeochemical parameters (e.g., Garnier et al. 2016).

In the ensemble OSSE framework introduced here, the input of the problem becomes a probability distribution of possible “true” states (describing the prior uncertainty), and so the original “truth” is no longer directly available or relevant. Each ensemble member could thus be alternatively used as possible “true” state to simulate observations from candidate observing arrays. These observations can then be assimilated into the other available members (leaving out the member used as the “truth”), and produce updated ensemble members that can finally be used along with a suite of probabilistic skill scores (e.g., Toth et al. 2003; Candille et al. 2007, 2015) to assess the impact of the assimilated data using each “truth” as verification.

Here, this integrated ensemble-based probability score approach is applied to a single assimilation time step to demonstrate the benefits of implementing such an approach, and give some preliminary insights about possible deployment strategies of the future BGC-Argo network in the North Atlantic. We use a probabilistic version of a coupled physical–biogeochemical model (inherited from Garnier et al. 2016) to evaluate two distributions of BGC-Argo arrays, and their combined value with satellite ocean color data.

The following section aims to provide a short summary of the classical OSSE procedures, while section 3 explains conceptually how these procedures may be extended to ensemble OSSEs, along with the probabilistic verification tools used to rigorously validate the conducted experiments. This novel type of OSSE is applied to a biogeochemical probabilistic system in the following methodology section (section 4), including detailed information about the multivariate assimilation update scheme and the simulated pseudo-observations. Section 5 first presents the set of OSSEs defined within the AtlantOS community, while the following subsections describe the impact of each experiment, and finally give some insights for some future design studies of the BGC-Argo array. A last section will give more general concluding remarks along with some caveats and limitations.

2. Background

a. Classical OSSE methodology

The design and the evaluation of ocean OSSE systems follows the long and well-established procedures used in atmospheric OSSEs since the early 1980s (e.g., Atlas et al. 2015). Typically, four main steps are required to perform an OSSE: 1) use a free-running circulation model to produce a nature run (NR), assumed to

represent the “true” state of the system; 2) simulate synthetic observations (including realistic errors) from this NR according to defined observing scenarios; 3) incorporate these newly generated observations into a data-assimilative run, known as the control run, which is usually different from the NR (e.g., different model, initial/forcing conditions, parameterizations and resolution) to produce an updated run that reflects the observations; and 4) assess the performance (score) of each tested scenario, which relies on comparisons of the magnitude and distribution of root-mean-square (RMS) errors between the NR, the control run, and the updated run. Specific and rigorous guidelines were adopted in the meteorological community to avoid possible bias in error growth between the NR and the data-assimilative model simulation (e.g., Atlas et al. 1985a,b; Arnold and Dey 1986; Hoffman et al. 1990), including situations referred to as the “identical twin” or the “fraternal twin” problems. Running realistic OSSEs also requires a prior validation of the NR to ensure that key phenomena measured by the observing systems are reproduced with sufficient accuracy. Moreover, the various types of errors present in real observations (e.g., instrumental errors, calibration errors) need to be properly incorporated, as well as representativeness errors (i.e., differences coming from unresolved or poorly resolved phenomena in the NR model). Failure to correctly add the different errors will lead to overestimates or underestimates of the observing system impacts.

A rigorous assessment of an OSSE system also includes a first comparison with a reference OSE (e.g., Atlas 1997) to evaluate the impact of present-day observing systems. The OSSE system is validated if consistent observation impacts are found between the OSSE and the reference OSE. However, it is only recently that similar validation strategies were developed in ocean OSSE studies, including comparisons with reference OSEs (e.g., Halliwell et al. 2014).

In most cases, OSSE systems involve single NR and control run, using either standalone (e.g., ocean circulation models) models or coupled modeling systems (i.e., simulating simultaneously the evolution of two related components, such as the ocean circulation and marine ecosystems). However, recent recommendations describing future improvements in OSSE systems (Schiller et al. 2015; Hoffman and Atlas 2016) underline the growing interest of ensemble simulations to better quantify model uncertainties and to ensure realistic OSSE results.

b. Status of ensemble OSSE systems

Since the last two decades in NWP centers, ensemble-based Monte Carlo techniques are widely used to predict

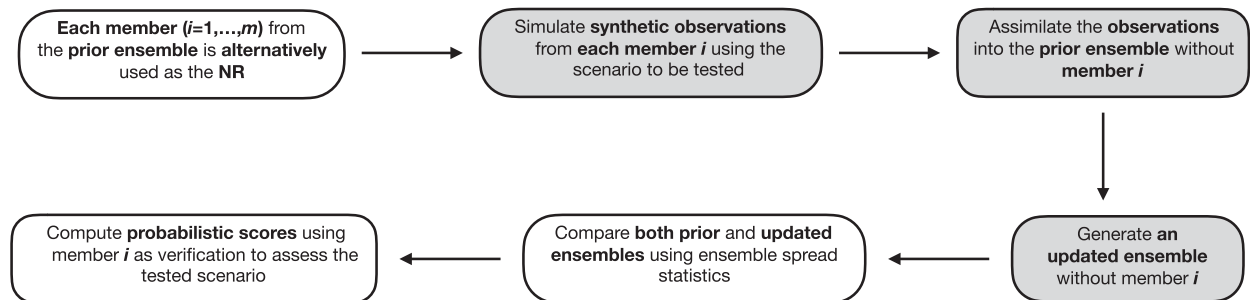


FIG. 1. Schematic of the cross-validation algorithm computed over an m -member ensemble. The data assimilation steps are shaded in gray.

future probability distributions of the state of the atmosphere. Stochastic parameterizations [see Leutbecher et al. (2017) for a review] were implemented to explicitly simulate various types of uncertainty (e.g., in the initial conditions and in the model's physical equations) and, therefore, to better reflect the chaotic nature of the atmosphere. To keep pace with operational systems, ensemble Kalman filter (EnKF) techniques (e.g., Evensen 2003) were adopted to perform OSSEs. For example, EnKF-based OSSEs were carried out to examine various Doppler radar networks in the lower atmosphere (Snyder and Zhang 2003; Tong and Xue 2005; Xue et al. 2006), or to test future altimeter configurations (Mourre et al. 2006; Le Hénaff et al. 2008, 2009).

In these OSSE studies, there is typically no comparison to any NR or control run after updating the ensemble simulation with synthetic observation data, the evaluation of the assimilation scheme is only based on temporal and spatial variations of the ensemble spread. The model uncertainty is characterized by the ensemble dispersion, and the impact of observations is generally based on RMS misfits (errors) between the ensemble without assimilation and the one with assimilation of synthetic data from each tested scenario.

3. Ensemble OSSE system design

OSSEs based on ensemble or hybrid data assimilation systems are thus generally evaluated using the similar metrics adopted for deterministic systems (i.e., using a single data-assimilative run), even though using ensembles give the opportunity to compare probability density functions (PDFs), generated by the prior ensemble members (before assimilation) and the updated ensemble members (after assimilation). This section presents a new framework to perform this upcoming probabilistic type of OSSEs, including appropriate verification tools for evaluating both existing and future observation systems.

a. Synopsis of the methodology: A cross-validation approach

Assuming we have an m -member ensemble that correctly describes the prior uncertainty (ensemble spread), each member can alternately be used as the NR using a cross-validation algorithm (Fig. 1). Two advantages are offered by this method: 1) the use of multivariate states rather than a single one, and 2) no subset of the prior ensemble has to be used as possible nature runs. The algorithm consists of looping on the m -member ensemble ($i = 1, \dots, m$) to perform the following steps: simulate synthetic observations from each member i used as the NR, solve the assimilation problem using the ensemble (leaving out member i), and then assess the quality of the updated ensemble using member i as verification, and the prior ensemble as the control run. The problem must thus be solved m times and can be computationally expensive depending on the ensemble size, which could be a limiting factor. Nonetheless, this approach can be applied in four dimensions (e.g., forecast context) despite demanding computational resources.

After accumulating enough realizations, an objective validation of the updated ensemble can be performed using probabilistic skill scores, typically decomposed into two properties: the reliability and the resolution [see Candille et al. (2015) for more detail]. The reliability checks the statistical consistency of the updated ensemble against the verifying member i . This is a necessary condition; however, reliability itself does not ensure a skillful probabilistic system, the resolution property is also required. In the context of OSSEs, the resolution can be seen as the system's ability to provide additional information (i.e., to reduce the uncertainty) after assimilating the synthetic observations from different candidate observing networks, or put in other words, the actual performance of the tested observing scenario. Practical ways to determine the reliability and the resolution are given below.

b. Probabilistic validation: An overview

The reliability can be graphically checked using rank histograms (see review by Hamill 2001). For scalar variables (e.g., sea level pressure or sea surface temperature), this skill score provides a graphic examination of the updated ensemble values with respect to the verifying data (in the present case from member i used as the NR and verification). The underlying principle is to rank the verification in the m -ensemble member values at each grid point. This process is repeated over all available realizations, considered independents to build a histogram over the possible ranks. The updated ensemble is considered to be reliable when it shows a flat rank histogram (i.e., a uniform distribution). An underdispersive updated ensemble will exhibit a U shape, meaning that a lack of variability in the ensemble underpopulates the middle ranks. In contrast, when the ensemble member values are from a distribution with an excess of variability, a bell shape results and the updated ensemble is called overdispersive. Additionally, a positive (or negative) bias in the update ensemble excessively populates the left (or the right) side of the rank histogram. However, the rank histograms do not allow evaluating the resolution, other tools are required to have a further assessment of any probabilistic system.

The continuous rank probability score (CRPS; Stanski et al. 1989; Hersbach 2000) is often used to provide a global skill assessment as it evaluates both the reliability and the resolution (e.g., Candille et al. 2007, 2015). For a given scalar variable, the CRPS is based on the square difference between the cumulative distribution function (CDF) of the updated ensemble and the CDF of the verifying member. The CRPS is similar to the mean absolute error typically used in deterministic OSSEs, and it has the dimension of the considered variables. In practice, the CRPS is integrated over an area and is typically averaged over the ensemble size, and therefore, it does not show the local impact of observations per se obtained by issuing ensemble OSSEs. This issue with the CRPS can make it difficult to discriminate different tested observing scenarios, leading the community to consider other types of verification tools for the resolution property.

Some studies (e.g., Roulston and Smith 2002; Bröcker and Smith 2007; Benedetti 2010; Peirolo 2011) suggest using information theory, and its concept of entropy to avoid condensing all the information gained by assimilating observations into a single numerical value. To this end, a probabilistic score based on entropy is introduced below.

c. Information theory: An alternative framework for a skill score

Let us first consider two PDFs, defined by two vectors \mathbf{p} and \mathbf{q} , assuming that \mathbf{p} is the “true” distribution of a defined variable (e.g., sea level pressure or sea surface temperature) and \mathbf{q} is the forecast PDF of this variable generated by an ensemble. One may use information theory to assess the quality of this forecast PDF. Nonetheless, prior to the definition of any information-based probabilistic score, three quantities need to be introduced, the entropy, the relative entropy and the cross entropy. The main concern here is only to present basic definitions of information theory, further detail can be found in Cover and Thomas (2012). The concept of information entropy has been introduced by Shannon (1948) to quantify how much information (uncertainty) is produced by a stochastic process and is given by

$$H(\mathbf{p}) = -\sum_{i=1}^n p_i \log_2 p_i, \quad (1)$$

where $H(\mathbf{p})$ is the information entropy, that is, the minimum possible number of bits required to encode the i th occurring event (where $i = 1, \dots, n$) from \mathbf{p} . The relative entropy $D(\mathbf{p}|\mathbf{q})$ is a typical measure of the distance between \mathbf{p} and \mathbf{q} , given by

$$D(\mathbf{p}|\mathbf{q}) = \sum_{i=1}^n (p_i \log_2 p_i - p_i \log_2 q_i). \quad (2)$$

As stated by Cover and Thomas (2012), the relative entropy (also known as the Kullback–Leibler divergence) is not a true distance between \mathbf{p} and \mathbf{q} ; it can rather be interpreted as the number of extra bits required to encode on the average any events drawn from the distribution \mathbf{q} rather than \mathbf{p} (considering that \mathbf{p} is the “true” distribution). The average number of total bits assigned to the events distributed by \mathbf{q} rather than \mathbf{p} is defined as

$$H(\mathbf{p}, \mathbf{q}) = H(\mathbf{p}) + D(\mathbf{p}|\mathbf{q}), \quad (3)$$

where $H(\mathbf{p}, \mathbf{q})$ is known as the cross entropy. Since entropy corresponds to the minimum of encoded bits, note that $H(\mathbf{p}, \mathbf{q}) > H(\mathbf{p})$ [except if $\mathbf{q} = \mathbf{p}$, then $H(\mathbf{p}, \mathbf{q}) = H(\mathbf{p})$]. Combining (1) and (2), cross entropy can be written as

$$H(\mathbf{p}, \mathbf{q}) = -\sum_{i=1}^n p_i \log_2 q_i. \quad (4)$$

Roulston and Smith (2002) defined $H(\mathbf{p}, \mathbf{q})$ as the expected value of ignorance (IGN), also referred to as information deficit, a skill score used in their study

for the evaluation of probabilistic forecasts that were compared against a nature run. [Roulston and Smith \(2002\)](#) pointed out that a single minimum of IGN could be found if and only if the forecast PDF \mathbf{q} coincides with the true PDF drawn by the nature run \mathbf{p} (i.e., when $\mathbf{q} = \mathbf{p}$ and $D(\mathbf{p}|\mathbf{q})$ is null). Based on this condition, [Roulston and Smith \(2002\)](#) suggested that on average over a large set of forecasts, the expected ignorance can be interpreted as the entropy of the forecast $H(\mathbf{q})$ itself.

To assess the effective impact of assimilating observations within the OSSE framework, one may normalize the expected ignorance score (e.g., on a common scale from 0 to 1) to facilitate comparisons between the different tested observing scenarios, as IGN values may vary depending on the shape and/or the sample size of the considered PDF. For instance, if the probabilistic distribution \mathbf{q} considered to represent the true distribution \mathbf{p} is uniform (i.e., the outcome of any event is equally probable), the entropy of a random variable χ taking $i = 1, \dots, n$ values is maximized by $q_i = 1/n$, and a maximum value of entropy H_{\max} is given by $\log_2 n$ ([Shannon 1948](#)). However, if a nonuniform probability distribution \mathbf{q} is considered to represent \mathbf{p} , $q_i \neq 1/n$, and thus, $H_{\max} < \log_2 n$.

Assuming that, over a large set of independent realizations, the average entropy is a good estimate of ignorance, a normalized IGN skill score (hereinafter IGN_n) defined over the $[0, 1]$ interval can be computed as the ratio between entropy $H(\mathbf{p})$ and cross-entropy $H(\mathbf{p}, \mathbf{q})$, since $H(\mathbf{p}, \mathbf{q}) \geq H(\mathbf{p})$:

$$\text{IGN}_n = \frac{H(\mathbf{p})}{H(\mathbf{p}, \mathbf{q})}. \quad (5)$$

In what follows, a simple example is presented to connect the theory in this subsection to the results later. Consider a PDF of chlorophyll produced by an ensemble simulation for each grid point over the North Atlantic, and the binary event of whether the chlorophyll will be below or above the observed seasonal mean. In that simple case, the initial PDF \mathbf{p} is the chlorophyll PDF for each geographical location, while the distribution \mathbf{q} is just a threshold value corresponding to the seasonal mean at each location. A map of the event's outcomes can be constructed by counting the number of ensemble members that are below or above the mean value. Ignorance can then be used to measure how well the ensemble agrees with the seasonal mean.

d. Summary: A verification package to evaluate ensemble OSSEs

How to produce probabilistic OSSEs, and how to evaluate them is our focus in this paper. A cross-validation

algorithm was proposed to take into account multi-nature runs instead of one, and use a suite of verification tools that enable a comprehensive evaluation of the information brought by the observations based on two properties of any probabilistic system, the reliability and the resolution. The reliability can first be assessed using the rank histograms, while a global evaluation of each OSSE performance can then be achieved using the CRPS, as it provides a condensed evaluation of the system's reliability and resolution. To examine the spatial distribution of the observation impact, one can further investigate the resolution property by using the normalized ignorance skill score presented above.

4. Experiment design: A biogeochemical application

For marine biogeochemistry, several recent studies ([Dowd 2011](#); [Doron et al. 2011, 2013](#); [Fontana et al. 2013](#); [Garnier et al. 2016](#)) have made use of using a stochastic-like formulation to correct model uncertainties, which can play a key role in estimating the dynamical behavior of marine ecosystems. The effect of these uncertainties are mostly the result of nonlinearities in the model equations and various biogeochemical model imperfections (e.g., simplified biology, unresolved biological diversity, unresolved scales).

The recent ensemble simulation from [Garnier et al. \(2016\)](#) used stochastic processes to explicitly simulate the joint effects of uncertain biological parameters and unresolved scales into a coupled physical–biogeochemical model in a $1/4^\circ$ North Atlantic configuration. The ensemble was able to simulate consistent surface chlorophyll distributions with satellite ocean color data (SeaWiFS) over the North Atlantic basin. Only relevant features of this ensemble simulation (hereinafter the prior ensemble) are presented below, while a thorough description along with the model configuration can be found in [Garnier et al. \(2016\)](#).

As part of AtlantOS and building on the experience inherited from this study, a set of biogeochemical ensemble OSSEs (see next section) has been performed to investigate the impact on the prior ensemble of assimilating synthetic observations from two possible BGC-Argo array distributions, including their combination with satellite ocean color data. Nevertheless, as we are using a large (60 member) ensemble simulation, it was decided to restrict this application to a single assimilation time step (i.e., no feed forward impacts), in order to reduce the numerical cost of the biogeochemical OSSEs.

Our approach here can thus be seen as a showcase rather than as a thorough assessment of the tested observing scenarios, and so one needs to keep in mind that any guidelines resulting from these OSSEs can only be considered as preliminary insights for future assimilation experiments.

a. The probabilistic coupled physical–biogeochemical model

The physical component of the model is based on the NEMO/OPA code (Madec 2008) implemented in the North Atlantic Ocean at $1/4^\circ$ horizontal resolution, including 46 vertical levels (a DRAKKAR configuration called NATL025; Barnier et al. 2006). The model is forced by the ERA-Interim ECMWF atmospheric fields (Uppala et al. 2005). NATL025 was initialized with the Levitus climatology (Levitus et al. 1998) to generate a 13-yr physical model spinup. The biogeochemical component of the coupled ensemble simulation is the PISCES-v2 (Aumont et al. 2015) model at $1/4^\circ$ horizontal resolution, covering the NATL025 domain from 20°S to 80°N and from 98°W to 23°E . PISCES-v2 contains 24 prognostic biogeochemical variables that are advected and diffused in three-dimensional space and at each time step by the physical model. The regional $1/4^\circ$ PISCES-v2 model was initialized in January 2002 from a global $1/4^\circ$ PISCES-v2 simulation to generate a biogeochemical spinup of 3 years between January 2002 and December 2004. Note that all members have same atmospheric forcing, and the physical ocean components (i.e., u , v , temperature, and salinity) do not vary with ensemble member. Only key biogeochemical variables with a direct impact on primary production vary.

The prior ensemble described in Garnier et al. (2016), includes 60 members over a 1-yr period based on direct stochastic parameterizations (following Brankart et al. 2015) of two classes of biogeochemical uncertainties, resulting from approximated biogeochemical parameters and unresolved scales. The stochastic parameterizations are uniformly implemented over the water column, though necessarily the simplest, this approach is a realistic hypothesis as each ensemble member is able to simulate coherent vertical distributions with various behaviors at a given grid point between the different members. Despite a slight underdispersion, the surface chlorophyll patterns simulated by the prior ensemble were found to be consistent with SeaWiFS observations for three dates during 2005 that exhibit different biological activity features. Below the surface, the vertical structure of chlorophyll was correctly represented over the euphotic layer (0–200 m) and appeared to be strongly correlated with the surface distribution.

As explained above, the cross-validation algorithm is only applied to a three-dimensional (3D) assimilation problem, that is, for 15 April 2005, which is roughly a month before the spring bloom period identified during May–June 2005. Overall, a good agreement is found between SeaWiFS data (not shown) and the relatively low surface chlorophyll concentrations described by the ensemble members (see statistics in Fig. 2). Both the ensemble and the observations show higher concentrations at latitudes between 30° and 50°N (especially along the coasts), as well as an elongated structure of lower chlorophyll centered on 20°N . Among the ensemble members, most of the chlorophyll dispersion is observed along the Gulf Stream pathway and coastal areas (see Fig. 2d), in addition to significant differences observed within the subtropical gyre. Inversely, the chlorophyll dispersion is small in the Mediterranean Sea and at high latitudes above 50°N (e.g., south of Greenland and in the Labrador Sea). Similar dispersion is shown at the subsurface down to 50 m deep, while below, lower ensemble dispersion in chlorophyll is generally observed.

Within our application's framework, the ensemble simulation for 15 April 2005 describes the multivariate biogeochemical states that will be used to perform our probabilistic OSSEs, for a single time step assimilation, and so the evaluation of these experiments is limited to the impact on spatial patterns.

b. The data assimilation method

The assimilation update scheme is based on a localized version of a square root algorithm in ensemble Kalman filters (e.g., Bishop et al. 2001; Evensen 2003). An eigenbasis algorithm [see Brankart et al. (2010) for more detail] is used to efficiently generate the observational update, based on the singular evolutive extended Kalman (SEEK) proposed by Pham et al. (1998). The i th prior ensemble member (where $i = 1, \dots, m$ and $m = 60$) is individually updated and decomposed as

$$x_i^{\text{pr}} = \bar{x}^{\text{pr}} + \delta x_i^{\text{pr}}, \quad (6)$$

where \bar{x}^{pr} is the prior ensemble mean and δx_i^{pr} the corresponding anomalies, which are written in square root form to compute the observational update using the eigenbasis algorithm, as described in Brankart et al. (2010). For this purpose, the following eigenbasis decomposition of the matrix Γ is computed:

$$\Gamma = (\mathbf{H}\mathbf{S}^{\text{pr}})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{S}^{\text{pr}}) = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}, \quad (7)$$

where $(\mathbf{H}\mathbf{S}^{\text{pr}})$ is the square root covariance of the prior ensemble matrix in the observation space, and \mathbf{R} is the

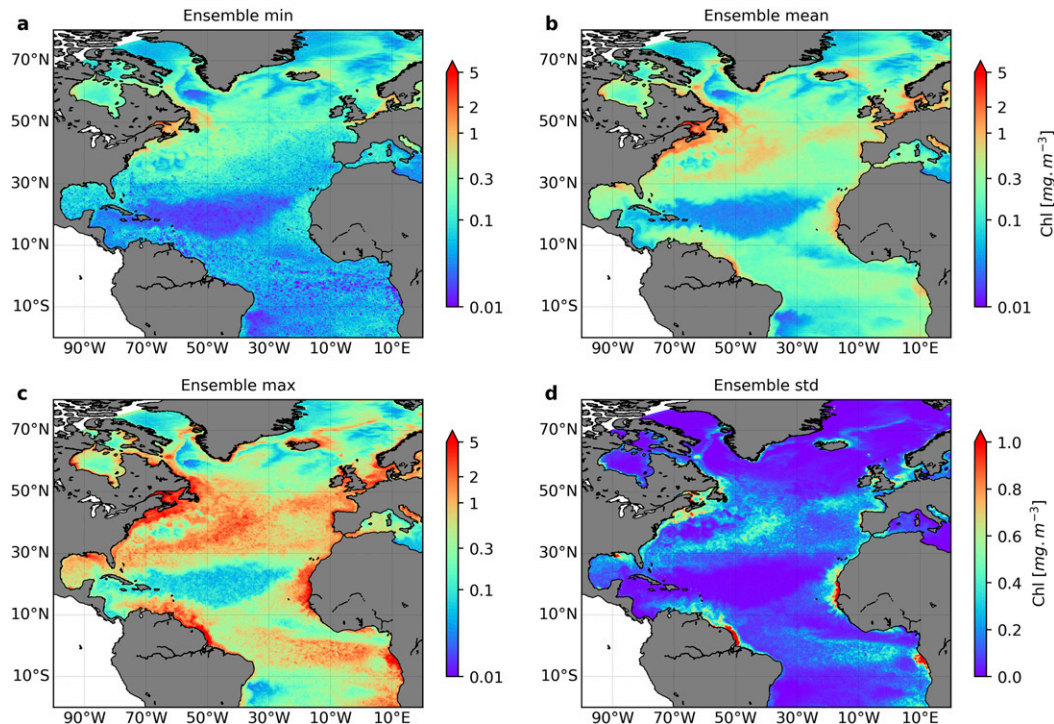


FIG. 2. Statistics of the surface chlorophyll distribution simulated by prior ensemble simulation for 15 Apr 2005. Surface chlorophyll ensemble (a) minimum and (c) maximum. (b) The surface chlorophyll ensemble mean and (d) the standard deviation (std). The color bar is in \log_{10} scale in (a)–(c).

observation error covariance matrix. The eigenvalues and the eigenvectors of $\mathbf{\Gamma}$ are provided by $\mathbf{\Lambda}$ (diagonal matrix), and \mathbf{U} (the unitary matrix). Note that \mathbf{R} is a diagonal matrix in which entries inside the main diagonal are the observation error standard deviation associated with each synthetic observation. This value takes into account both instrumental and representativity errors, and is set to 30% of the chlorophyll concentration (see next section) to be consistent with previous studies that only assimilated ocean color data (e.g., [Ciavatta et al. 2011](#); [Fontana et al. 2013](#); [Ford and Barciela 2017](#)).

The updated ensemble mean \overline{x}^{up} is then defined using the matrix $\mathbf{\Gamma}$ as

$$\overline{x}^{\text{up}} = \overline{x}^{\text{pr}} + \mathbf{S}^{\text{pr}} \mathbf{U} (\mathbf{I} + \mathbf{\Lambda})^{-1} \mathbf{U}^T (\mathbf{H} \mathbf{S}^{\text{pr}})^T \mathbf{R}^{-1} (\mathbf{y}_0 - \overline{\mathbf{H}} \overline{x}^{\text{pr}}), \quad (8)$$

while each i th updated ensemble anomaly is defined as

$$\delta x_i^{\text{up}} = \sqrt{m-1} [\mathbf{S}^{\text{pr}} \mathbf{U} (\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{\Lambda}^{-1/2} \mathbf{U}^T]_i. \quad (9)$$

Now, each updated member δx_i^{up} can be rewritten as the sum of the updated ensemble mean and the updated anomalies:

$$x_i^{\text{up}} = \overline{x}^{\text{up}} + \delta x_i^{\text{up}}. \quad (10)$$

Furthermore, considering the nonlinear relationships between the different biogeochemical state PISCES-v2 variables and associated parameters, the useful Gaussian assumption (i.e., which allows linear transformations to solve the observational update problem) is not expected to hold. Therefore, anamorphosis transformations (e.g., [Bertino et al. 2003](#); [Béal et al. 2010](#)) are applied to each separate variable of the state vector prior to the ensemble update to ensure that the marginal PDF of each variable becomes close to Gaussian, as it has been previously done in related studies (e.g., [Doron et al. 2011, 2013](#); [Fontana et al. 2013](#)). With anamorphosis, the specification of the observation errors needs to take into account the nonlinear transformation applied to state variables, and so the error associated with each observation is set following [Brankart et al. \(2012\)](#), that is, by multiplying the observation error standard deviation by the local slope of the anamorphosis transformation. The inverse local anamorphosis transformations are performed after assimilation to return to the original model space. The anamorphosis presents two advantages: 1) a better description of the relationship between observed and nonobserved variables and 2) a

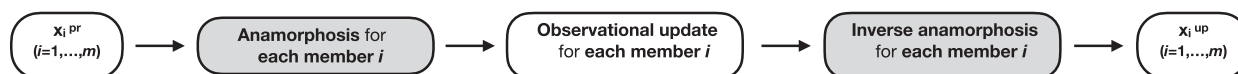


FIG. 3. Schematic describing how local anamorphosis transformations (gray shading) are applied to each model variable over the 60-member PISCES probabilistic simulation.

parameterization of the error statistics that avoids obtaining negative values for the concentration variables of the state vector after the assimilation step. For the sake of clarity, Fig. 3 presents the main steps of how anamorphosis is applied within the assimilation system. While, the anamorphosis approach has been shown to provide an effective way of describing uncertainties in coupled physical–biogeochemical models (e.g., Doron et al. 2011; Brankart et al. 2012; Fontana et al. 2013), a possible alternative to the anamorphosis algorithm applied in this paper would be to use the gamma, inverse-gamma, and Gaussian EnKF (GIGG-EnKF) developed by Bishop (2016). Even though the GIGG-EnKF is more suited for variables whose uncertainty is well represented by gamma and/or inverse-gamma distributions, it may also be appropriate to deal with biogeochemical variables (e.g., chlorophyll concentrations), and thus be applicable to ocean biogeochemical assimilation systems.

A localization algorithm (Brankart et al. 2011) is also used to avoid unrealistic effects of large spatial correlations. The assimilation of the synthetic observations is performed locally, limited by a radius of influence set to one grid point and the cutoff radius (i.e., the distance at which the weight of the observations is negligible) to three grid points. These two values are based on various assimilation experiments that aimed to determine a noticeable spread reduction of the updated ensemble without degrading the probabilistic reliability property.

The System of Sequential Assimilation Modules (SESAM) software [see Brankart et al. (2012) for further detail] was used to compute all matrix operations required by the assimilation scheme, such as the innovation vector $\mathbf{y}_0 - \mathbf{H}\mathbf{x}^{\text{pr}}$, the assimilation update and associated covariance errors. The state vector included all prognostic biogeochemical state variables of PISCES-v2 (no dynamics as mentioned above), meaning that the assimilation update was multivariate.

c. Synthetic chlorophyll observations

The assimilated datasets include chlorophyll concentrations simulated at the locations of real satellite ocean color data observed for 1 January 2009 and daily Argo float trajectories generated as part of AtlantOS. Ocean color observations were simulated based on the actual data coverage provided by CMEMS, that is, the global

level three daily merged product gridded at a spatial resolution of 4 km on an sinusoidal grid (detailed information is given in the product user manual, available online at <http://marine.copernicus.eu/documents/PUM/CMEMS-OC-PUM-009-ALL.pdf>). The derived chlorophyll product that is used in this study has been generated by ACRI-ST (<http://hermes.acri.fr>) using the Copernicus-GlobColour processor (see Maritorea et al. 2010), and based on the three sensors available in 2009 [MODIS *Aqua*, SeaWiFS, and Medium Resolution Imaging Spectrometer (MERIS)]. The daily ocean color data coverage is similar between 2009 and present day, though the current dataset is originating from the three merged sensors MODIS *Aqua*, Ocean and Land Colour Instrument (OLCI), and VIIRS. It is noteworthy that selecting the ocean color data coverage from 1 January 2009 was not an arbitrary choice; it allows us to compare contrasted regions (whether ocean color is available or not) in “test mode” before conducting future OSSEs over the year 2009, and so to keep pace with the companion study led by the Met Office that performed deterministic OSSEs in “operational mode” from 1 January to 31 December 2009.

The BGC-Argo trajectories are based on a quasi-homogeneous global distribution (see Gasparin et al. 2019) with around one profile per $3^\circ \times 3^\circ$ box per 10 days over the 2009–11 period. To avoid under-sampling in the tropics and in the South Atlantic region, some artificial float trajectories were added to the dataset of 2009 from the float profiles deployed in 2010 prior to performing the AtlantOS physical OSSEs. Here, a first BGC-Argo distribution was considered by aggregating those float positions over a full cycle of 10 days (i.e., from 1 to 10 January 2009). In practice, it corresponds to having biogeochemical profiles at the same spatial and temporal resolution as *T/S* profiles recorded by the actual Argo array (~ 4000 floats). A second BGC-Argo distribution was also considered by aggregating this time the float trajectories over 3 days (1–3 January 2009), representing about a quarter of the existing Argo floats (~ 1000 floats).

As we mentioned above, the observation error associated with the chlorophyll concentration was set to 30%, for the satellite ocean color and the two BGC-Argo arrays, a value used in other studies (e.g., Fontana et al. 2013; Ciavatta et al. 2018). The data coverage of the two BGC-Argo distributions and the daily

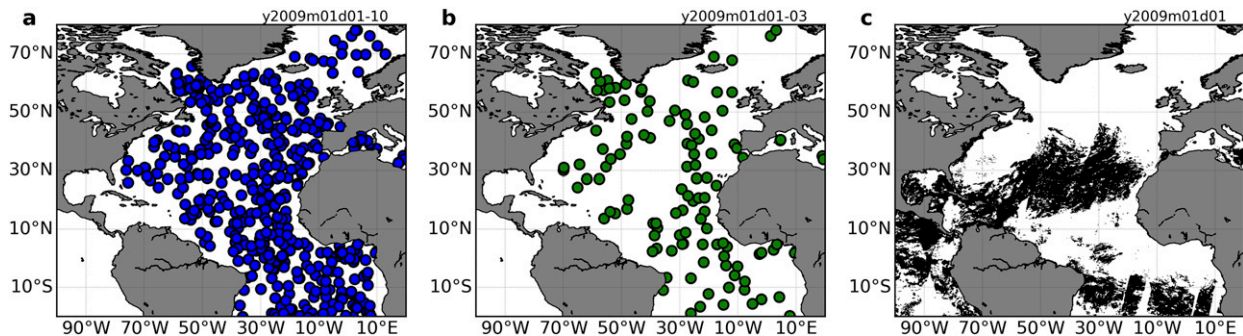


FIG. 4. Assimilated observation networks to assess the defined scenarios at dates as indicated in legends (see upper-right corner of each panel). (a) Blue dots indicate a quasi-homogeneous Argo distribution, around one profile per $3^\circ \times 3^\circ$ box per 10 days; (b) green dots indicate a 1/4 subsample of this Argo array, and (c) daily ocean color tracks extracted from the Copernicus Marine Environment Monitoring Service (CMEMS) database.

ocean color observations for 1 January 2009 are presented in Fig. 4.

5. Experiments

We present in this section a series of four basic experiments assessing the impact of chlorophyll observations on the prior ensemble simulation, including some preliminary recommendations for the assimilation of BGC-Argo data in the North Atlantic, and further perspectives of development.

a. Scenarios under consideration

Within the AtlantOS initiative, dedicated OSSEs are performed to assess the value of the future extension of Argo to biogeochemical variables (see Johnson and Claustre 2016a) in close synergy with existing satellite ocean color data (which are only effective near the sea surface and in cloud-free conditions). It is expected that such a system would enable an unprecedented comprehensive view of the interactions between climate and marine ecosystems (e.g., variability in biological productivity, ocean uptake of CO_2 , or ocean acidification).

The focus here is set on assessing the two distributions of BGC-Argo floats described above (see experiments A and B in Table 1). The first distribution was chosen because it represents the target number of BGC-Argo floats (Johnson and Claustre 2016b), while the second distribution was selected to be the closest of the existing Argo array to assess the value of having biogeochemical sensors on all floats. The floats are considered to have chlorophyll, nitrate, and oxygen sensors, though our main concern here is to only assess the impact of assimilating synthetic chlorophyll observations. Two additional experiments that combine the two BGC-Argo arrays and the daily CMEMS ocean color data coverage (experiments C and D in Table 1) were also

performed to assess the benefits of adding BGC-Argo arrays to the satellite ocean color system.

b. Impact of the observing scenarios

The four experiments presented above are evaluated using a classical deterministic score that relies on RMS errors, followed by probabilistic diagnostics using the rank histogram technique and the ignorance skill score (i.e., an information theoretic measure based on entropy) to evaluate the reliability and the resolution, respectively. The scoring results associated with the CRPS are however not shown, as it only provides a single number summary for both skill score properties that do not reflect the local impact of the synthetic chlorophyll observations after the assimilation step, and making it difficult to discriminate the four conducted experiments. In addition, we found that heterogeneous chlorophyll patterns with concentrations of different order of magnitude were mixed within each area used to compute the CRPS, which could give a misleading skill score for each experiment.

1) RMS ERROR METRIC

The impact of chlorophyll observations for each tested scenario can be assessed using a RMSE-based

TABLE 1. List of experiments performed to evaluate basic BGC-Argo future deployments.

| Experiments | Data assimilated and configuration |
|-------------|---|
| A | BGC-Argo on 1/4 of the nominal Argo array (~1000 floats) |
| B | BGC-Argo on the full nominal Argo array (~4000 floats) |
| C | Daily satellite ocean color data and BGC-Argo on 1/4 of the nominal array |
| D | Daily satellite ocean color data and BGC-Argo on the nominal array |

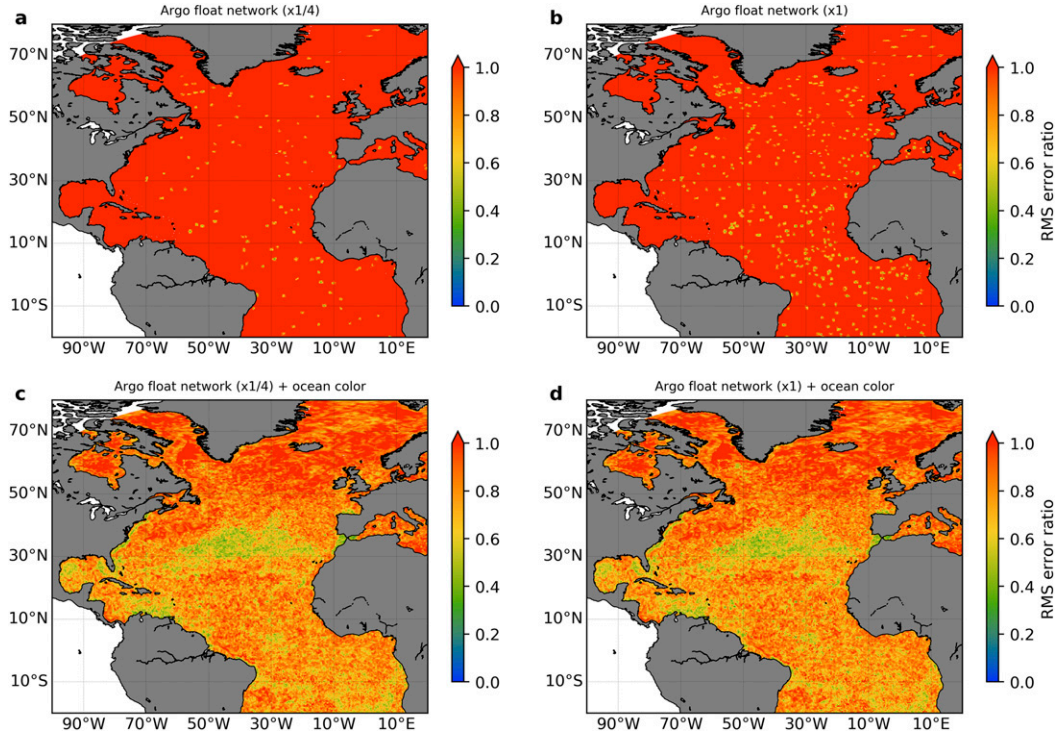


FIG. 5. Ratio of mean surface chlorophyll RMS errors between the updated ensemble and the prior ensemble for the defined scenarios. (a) BGC-Argo sensors on a quarter of the nominal Argo array, (b) BGC-Argo sensors on the full nominal Argo array, (c) daily satellite ocean color data and BGC-Argo on 1/4 of the nominal array, and (d) daily satellite ocean color data and BGC-Argo on the nominal array.

score (Fig. 5), as it is classically done with deterministic OSSEs, though as stated before there is no interpretation in terms of probability. For this purpose, the mean surface chlorophyll RMS errors between the ensemble (before and after assimilation) and each left out member used as the verification were first computed. The RMS error ratio (i.e., between the updated ensemble and the prior ensemble) was then calculated to give some insights about the observation system ability to reduce the prior uncertainty, and defined as

$$\text{RMSE}_r = \frac{\sum_{i=1}^m \sqrt{\frac{\sum_{j \neq i}^{m-1} (x_j^{\text{up}} - x_i)^2}{m-1}}}{\sum_{i=1}^m \sqrt{\frac{\sum_{j \neq i}^{m-1} (x_j^{\text{pr}} - x_i)^2}{m-1}}}, \quad (11)$$

where x_j^{up} and x_j^{pr} are, respectively, each ensemble member for the updated and the prior ensembles (with $j = 1, \dots, m-1$ and $m = 60$). The same verifying member is used before and after assimilating the synthetic observations, and is noted x_i (with $i = 1, \dots, m$).

The prior ensemble uncertainty related with surface chlorophyll is only reduced locally for the experiments A and B (Figs. 5a,b), where the synthetic observations of BGC-Argo floats are assimilated. In contrast, the two experiments combining both ocean color and BGC-Argo data (Figs. 5c,d) produce a widespread reduction (about 40%–50%) of the mean surface chlorophyll RMS error ratio (RMSE_r) over a zonal band across the Atlantic basin at around 30°N, as well as into the Gulf of Mexico and in the Caribbean Sea. Significant error reduction of about the same magnitude is also observed around 35°N, 10°W (near the Gibraltar Strait) and south of 10°S, where RMSE_r patterns are more patchy. As we could expect, the prior uncertainty associated with the surface chlorophyll is mostly reduced over the best satellite coverage, that is, where most of the synthetic ocean color data have been assimilated; however, significant changes are surprisingly not observed between experiments C and D.

Below the surface, RMSE_r values (not shown) for experiments A and B show only minor changes over the upper 200 m, while those for experiments C and D reveal complex vertical structures due to the assimilation of surface chlorophyll. This can be explained by the strong correlation between the surface chlorophyll

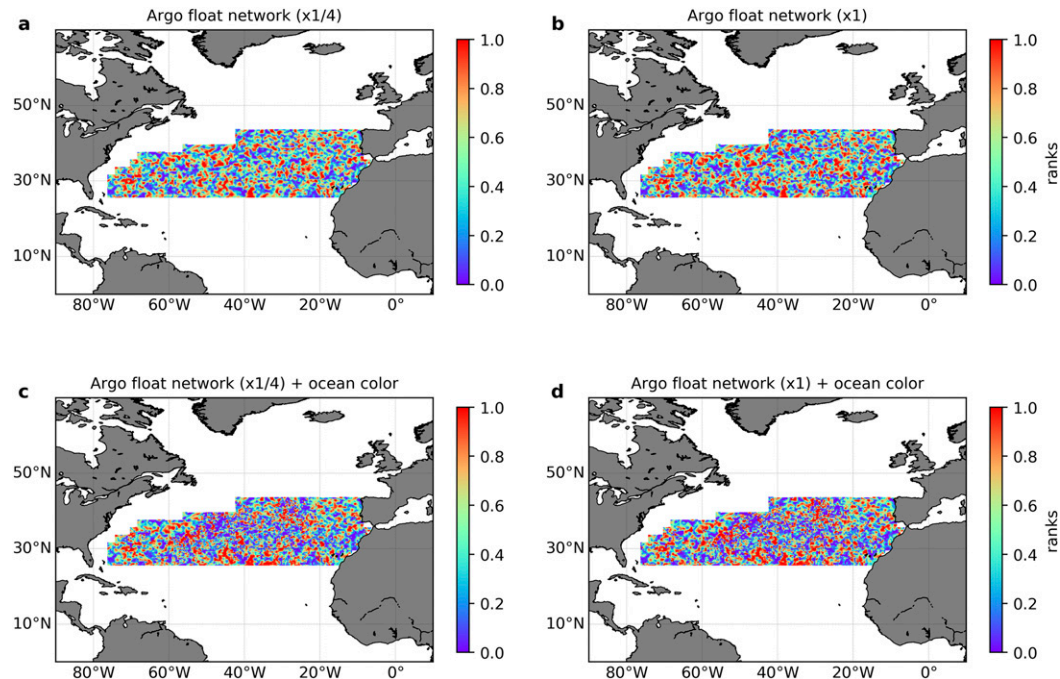


FIG. 6. Local rank distributions (see color bar) of surface chlorophyll over two Longhurst-defined ecological provinces located at the midlatitudes across the North Atlantic for the defined scenarios. (a) BGC-Argo sensors on a quarter of the nominal Argo array, (b) BGC-Argo sensors on the full nominal Argo array, (c) daily satellite ocean color data and BGC-Argo on 1/4 of the nominal array, and (d) daily satellite ocean color data and BGC-Argo on the nominal array.

distribution and the vertical chlorophyll patterns, as pointed out in section 4a. However, these RMSE_r vertical structures bring only little information regarding the reduction or not of the prior uncertainty along the vertical axis, and so we are not able to thoroughly assess the impact of assimilating BGC-Argo floats below the surface using this RMS error metric.

2) RANK SPATIAL DISTRIBUTIONS AND ASSOCIATED HISTOGRAMS

To take into account the various marine ecosystem's behaviors encountered across the globe, Longhurst (1995) proposed a comprehensive partition of the ocean into biogeochemical provinces (also known as regions) defined by both ocean dynamics and sea surface chlorophyll features. In Fig. 6, the local ranks are only accumulated at the surface over two biogeochemical regions across the Atlantic basin to avoid aggregating too heterogeneous chlorophyll patterns, such as those found near the coasts or over the Gulf Stream area. These two provinces have been defined in Longhurst (1995) as the North Atlantic subtropical west (NASW) and the North Atlantic subtropical east (NASE) provinces.

The resulting rank histograms are fairly flat for experiments A and B (Figs. 7a,b), suggesting that the chlorophyll distribution of the updated ensemble is

reliable (i.e., statistically consistent vis-à-vis the verification). However, a slight underdispersion (U-shape rank histograms) is identified for both experiments C and D (Figs. 7c,d), meaning that some verifications (about 20%) fall outside the ensemble after assimilation. Most of these outliers are found near 30°N between 60° and 20°W (Figs. 6c,d), where some extreme rank values (i.e., 0 or 1) can be observed. This might be due to a lack of variability (too small spread) within the prior ensemble or a slight bias in the assimilation process, where unrealistic surface chlorophyll correlation patterns could have been taken into account. Nevertheless, the rank histograms related to the ensemble members from experiments C and D are not far from being uniform, and so the reliability of surface chlorophyll distribution is considered to be verified. Note that adding ranks at greater depths does not significantly affect the rank histogram construction, while some changes in rank histogram shapes (although they remain nearly flat) can be found between the different Longhurst-defined provinces located in the North Atlantic Ocean (not shown).

3) ENTROPY-BASED SKILL SCORE

The main concern here is to present the impact of the chlorophyll observations using our modified ignorance

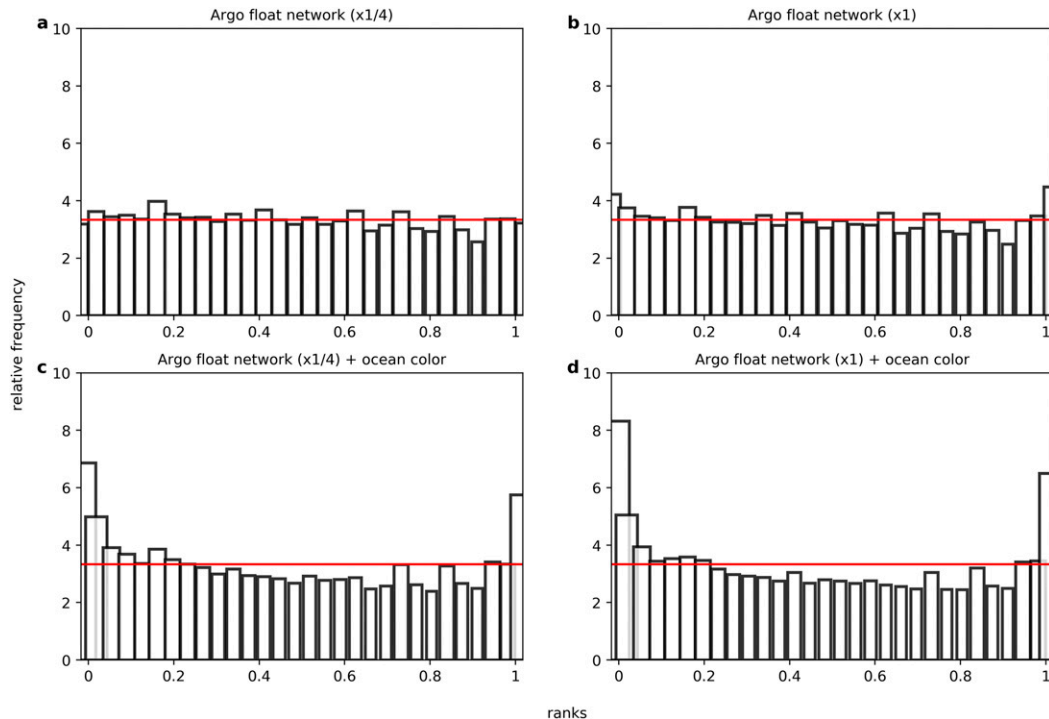


FIG. 7. Rank histograms of surface chlorophyll concentrations over the two Longhurst-defined ecological provinces located in the North Atlantic subtropical region for the defined scenarios. (a) BGC-Argo sensors on a quarter of the nominal Argo array, (b) BGC-Argo sensors on the full nominal Argo array, (c) daily satellite ocean color data and BGC-Argo on 1/4 of the nominal array, and (d) daily satellite ocean color data and BGC-Argo on the nominal array. For each grid point and all sorted members, the rank histogram indicates the frequency of occurrence (in percent) of the verifying value of chlorophyll. The red line indicates the ranks for a flat histogram.

skill score IGN_n . Following the example presented in the section 3c, we first need to identify a set of probabilistic events that are relevant to compare the different scenarios. IGN_n can then give a simple measure of the average information deficit with respect to the chosen occurring event. As a first attempt and for the sake of simplicity, let us consider the following binary event: “being below/above the median at the surface of the prior ensemble PDF.” In other words, this simple event aims to examine the spread of the chlorophyll distribution described by the ensemble after assimilation, compared to the values of surface chlorophyll concentration of the prior ensemble median. In that binary case, the probability distribution \mathbf{p} is uniform and when $\mathbf{p} = 0.5$, entropy $H(\mathbf{p})$ is at a maximum of 1 bit. In addition, cross-entropy $H(\mathbf{p}, \mathbf{q})$ equals $H(\mathbf{p})_{\max}$, as $\mathbf{q} = 1 - \mathbf{p} = 0.5$, and so entropy itself represents how much information deficit (uncertainty) is reduced after assimilation.

The surface entropy maps with respect to chlorophyll (Fig. 8) show, as expected, a reduction of prior uncertainty (i.e., entropy < 1 bit) where chlorophyll observations were assimilated. For experiments A and B

(Figs. 8a,b), the uncertainty is reduced locally at the positions of the synthetic BGC-Argo floats (entropy ranging from about 0.4 to 0.8 bit), which conforms with the $RMSE_r$ results. For experiments C and D (Figs. 8c,d), the uncertainty reduction mostly occurs over the best satellite data coverage, that is, at around 30°N across the basin, where $RMSE_r$ values suggested a spread contraction of the updated ensemble. Close inspection of the IGN_n values suggests a significant information gain in experiment D compared to experiment C, especially where more floats were added (e.g., in the equatorial region between 10°S and 10°N), whereas the corresponding $RMSE_r$ values showed little changes. Nevertheless, going toward lower IGN_n clearly suggests that information is added regarding the previous statement: “being below/above the median at the surface of the prior ensemble PDF,” while the $RMSE_r$ values have no particular meaning.

Near 50-m depth (near the maximum chlorophyll depth), knowledge with respect to the similar statement is mostly added by the two scenarios with the BGC-Argo arrays only (Figs. 9a,b), though significant gain is obtained over the 30°N latitudinal band and

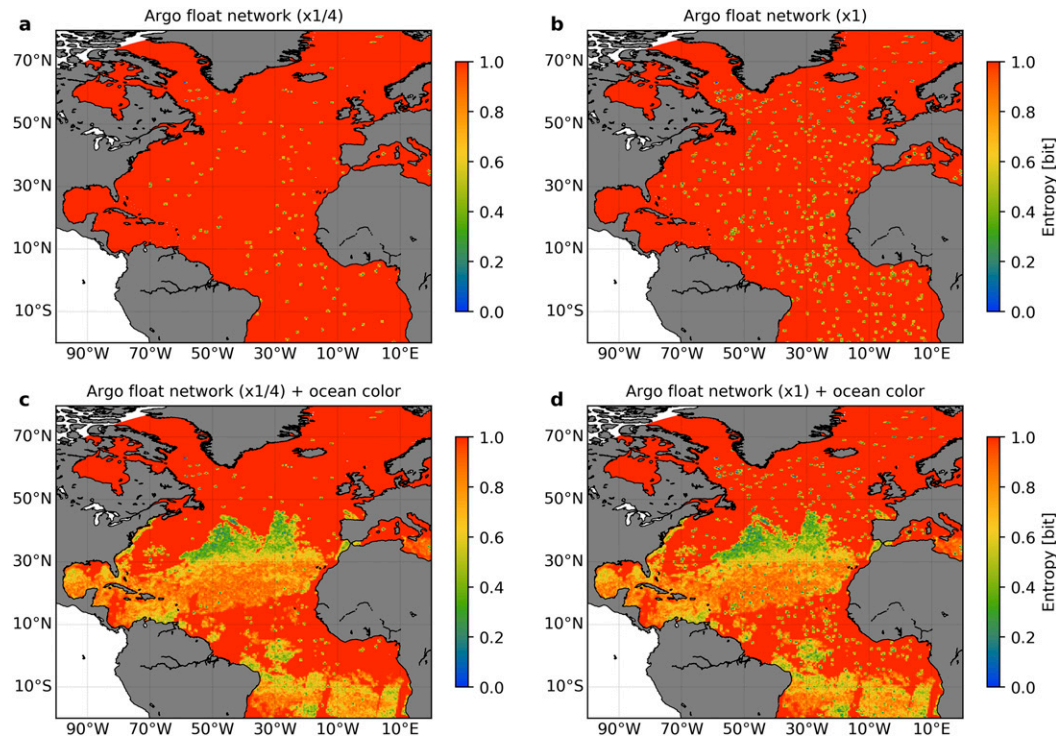


FIG. 8. IGN_n map of “being below/above the median of chlorophyll at the surface of the prior ensemble PDF” for its occurrence at the surface in the updated ensemble, and for the defined scenarios. (a) BGC-Argo sensors on a quarter of the nominal Argo array, (b) BGC-Argo sensors on the full nominal Argo array, (c) daily satellite ocean color data and BGC-Argo on 1/4 of the nominal array, and (d) daily satellite ocean color data and BGC-Argo on the nominal array. Red indicates the highest values of ignorance regarding the considered probabilistic event.

south of 10°S in experiments C and D (Figs. 9c,d). Note that the event “being below/above the median at the surface of the prior ensemble PDF” is not certain to occur in the updated ensemble at 50-m depth (unlike at the surface), and therefore, some areas in the maps where $IGN_n = 0$ are observed.

To further compare the different deployment scenarios, we examine a longitudinal section as a function of depth at 30°N. For the two first experiments (Figs. 10a,b), most of the impact is observed between 50- and 150-m depth, associated with evident vertical correlation structures. Nevertheless, a surprising result is that entropy exhibits quite similar patterns between the two BGC-Argo array distributions, suggesting that having an observing system with chlorophyll sensors on all existing Argo floats does not provide much more information (about the considered event) than having those sensors on about a quarter of the floats. Note that similar results are found for other longitudinal sections, for example at 5°N (see Fig. 11). However, further experiments that last longer than a day (e.g., a monthly period) will likely suggest some differences between the two BGC-Argo arrays.

As for experiments C and D (Figs. 10c,d), the prior uncertainty is further reduced at the surface down to 50–70-m depth, highlighting the impact of the satellite ocean color observations along the vertical axis. Although one may keep in mind that the strong correlation between the surface chlorophyll and its vertical distribution within the prior ensemble might lead to overestimated impact assessment of ocean color data over the uppermost euphotic layer.

Two other probabilistic events were also investigated to compare the four deployment scenarios. The statement “being inside/outside the quantile range 0.4–0.6 at the surface of the prior ensemble PDF” was considered to compute our ignorance skill score, and screen the updated ensemble PDF in relation to the surface chlorophyll concentrations distributed around the prior ensemble mean. To examine the updated ensemble PDF compared to the prior ensemble tails, we finally computed the IGN_n of “being inside/outside the quantile range 0.2–0.8 at the surface of the prior ensemble PDF.” Both cases exhibit similar spatial distributions of entropy compared to those obtained with the first probabilistic event, and so bring few additional information,

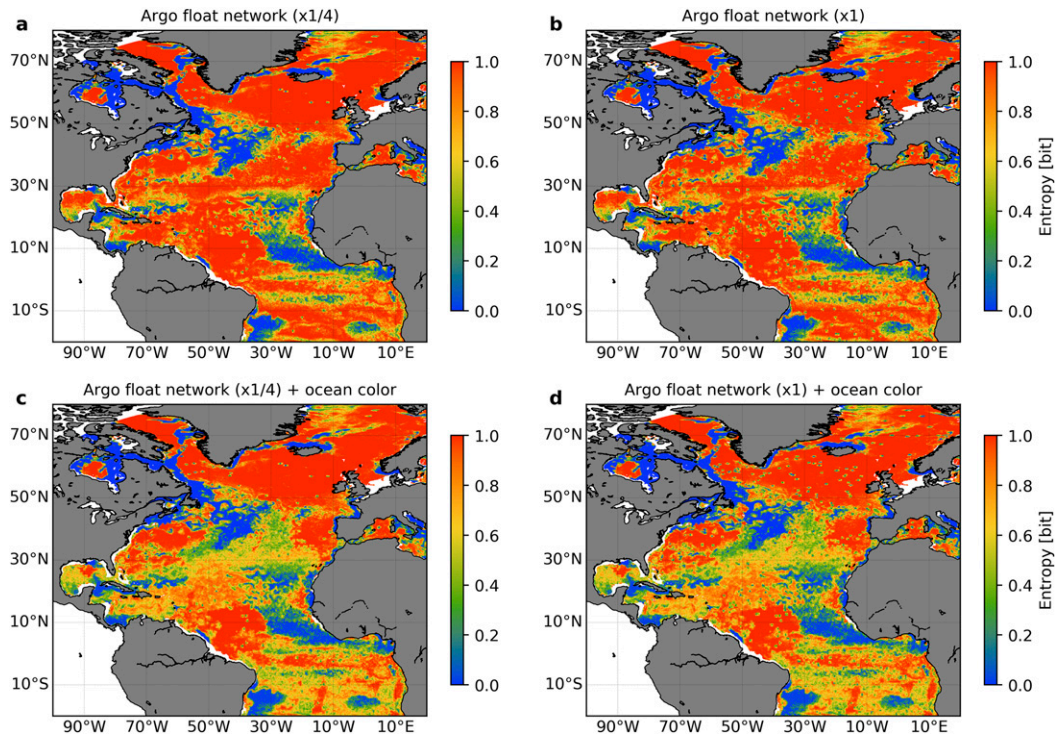


FIG. 9. IGN_n map of “being below/above the median of chlorophyll at the surface of the prior ensemble PDF” for its occurrence at 52-m depth in the updated ensemble, and for the defined scenarios. (a) BGC-Argo sensors on a quarter of the nominal Argo array, (b) BGC-Argo sensors on the full nominal Argo array, (c) daily satellite ocean color data and BGC-Argo on 1/4 of the nominal array, and (d) daily satellite ocean color data and BGC-Argo on the nominal array. Red and blue indicate, respectively, high and low ignorance, with respect to the considered probabilistic event.

though some differences in IGN_n values (i.e., the amount of bits) were noticed.

c. Assimilation of BGC-Argo data: Preliminary results and perspectives

The experiments presented above suggest that assimilating BGC-Argo floats significantly reduces the uncertainty associated with chlorophyll within the prior ensemble. A comparison of four basic deployment scenarios was first carried out using a classical metric relying on ratio of RMS misfits between the updated ensemble and the prior ensemble. The value of adding BGC-Argo to the actual satellite ocean color constellation was mostly observed where the satellite coverage is limited (i.e., at the northernmost latitudes and over the equatorial region between 10°S and 10°N). At the subsurface down to 150–200 m, the RMS error-like metric indicated strong vertical correlation structures, though it was not possible to make meaningful comparisons between the different scenarios. However, based on these first RMS error diagnostics, a straightforward recommendation, even if not new, is to deploy BGC-Argo floats with the highest sampling

frequencies in regions that are statistically more cloudy. Further experiments may thus investigate the float density required in those regions to complement in an optimal way the chlorophyll observations obtained by spatial ocean color sensors.

Regarding the probabilistic validation, we successfully assessed the statistical reliability of each conducted experiment using the rank histogram technique, as is done at NWP centers with ensemble forecasts. To investigate the actual impact of assimilating synthetic chlorophyll observations, we used a metric based on information entropy as previously done in few studies for probabilistic forecast schemes (e.g., [Roulston and Smith 2002](#); [Benedetti 2010](#); [Peirolo 2011](#)). For simplicity, we chose to look only at binary events such as “being below/above the median at the surface of the prior ensemble PDF,” although this entropy-based score can easily be extended beyond Bernoulli trials (i.e., experiments with more than two possible outcomes). At the surface, most of the information relative to the event selected above was gained, as expected, where the synthetic observations were assimilated. Below, the impact of satellite ocean color data is suggested

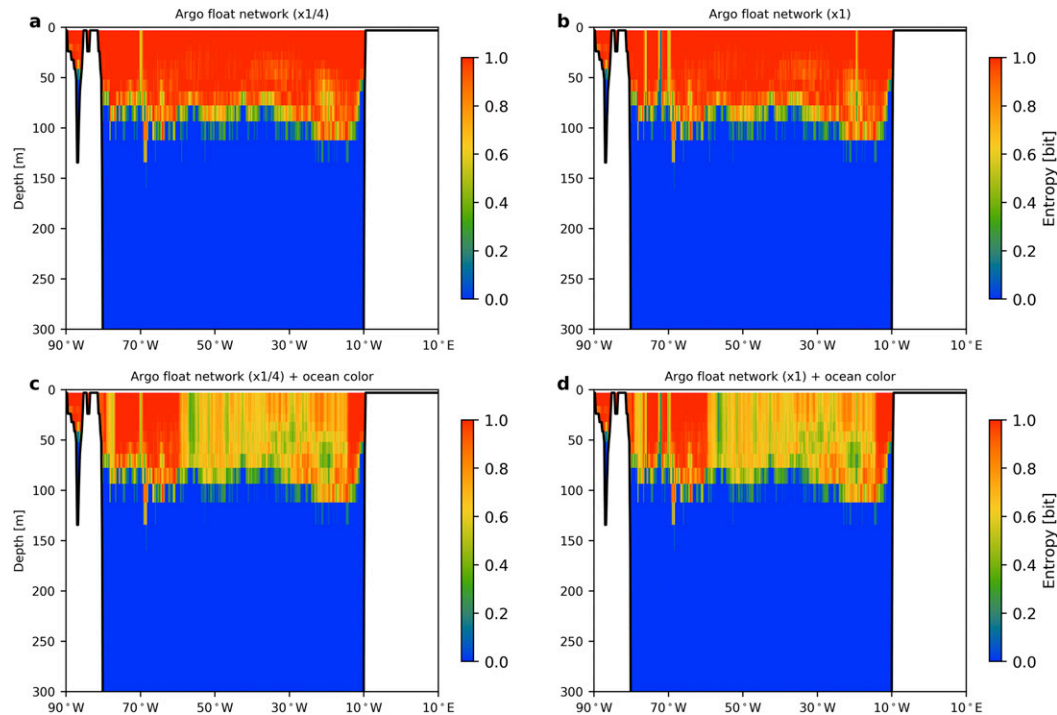


FIG. 10. Longitudinal section at 30°N of the chlorophyll IGN_r for the defined scenarios (color coded as in Fig. 9). (a) BGC-Argo sensors on a quarter of the nominal Argo array, (b) BGC-Argo sensors on the full nominal Argo array, (c) daily satellite ocean color data and BGC-Argo on 1/4 of the nominal array, and (d) daily satellite ocean color data and BGC-Argo on the nominal array.

to be confined over the top 50 m, while the chlorophyll observations from the two BGC-Argo arrays add information mostly over the 50–150-m depth range.

The preliminary conclusions that can be drawn from these ensemble-based OSSEs are 1) chlorophyll observations from the two BGC-Argo arrays provide valuable inputs in good synergy with ocean color data, especially where satellite information is limited such as over the equatorial region (consistent with the RMS error diagnostics), 2) assimilating BGC-Argo data lead to significant improvements at the subsurface, and 3) an array size of 1000 floats is a rational choice for the BGC-Argo network, as it significantly reduces the prior ensemble uncertainty. However, a realistic and effective evaluation of assimilating chlorophyll concentrations from both BGC-Argo array distributions must be performed in time as well (including periods encompassing spring algal blooms), but this is beyond the scope of this biogeochemical application, which merely aims to illustrate the generic ensemble-based OSSE approach presented in section 3. Note also that assimilating biogeochemical data is still challenging and immature; further developments of the current data-assimilation schemes may thus yield different results. Other sources of uncertainty should also be taken

into account to effectively assimilate chlorophyll observations. For example, uncertainties related to the physical ocean components (e.g., temperature and salinity) or uncertainties on other biogeochemical variables (e.g., the dissolved oxygen concentration, nitrates, or pH) may be introduced in the data-assimilation scheme.

6. Conclusions

In this paper, a generic cross-validation approach has been described to perform novel OSSE studies when an ensemble of data-assimilative simulation run is being used. Each ensemble member can be alternatively used as the “truth” to simulate synthetic observation data types (existing or not), while the other available members can be used to produce an updated ensemble that reflects the assimilated data. Two important advantages of this approach are 1) to provide an explicit description of model uncertainty to ensemble data assimilation systems, and 2) to allow objective statistical comparison between the prior ensemble system and the one updated by assimilation using a set of probabilistic verification skill scores similar to those routinely used in forecasting centers. Our approach

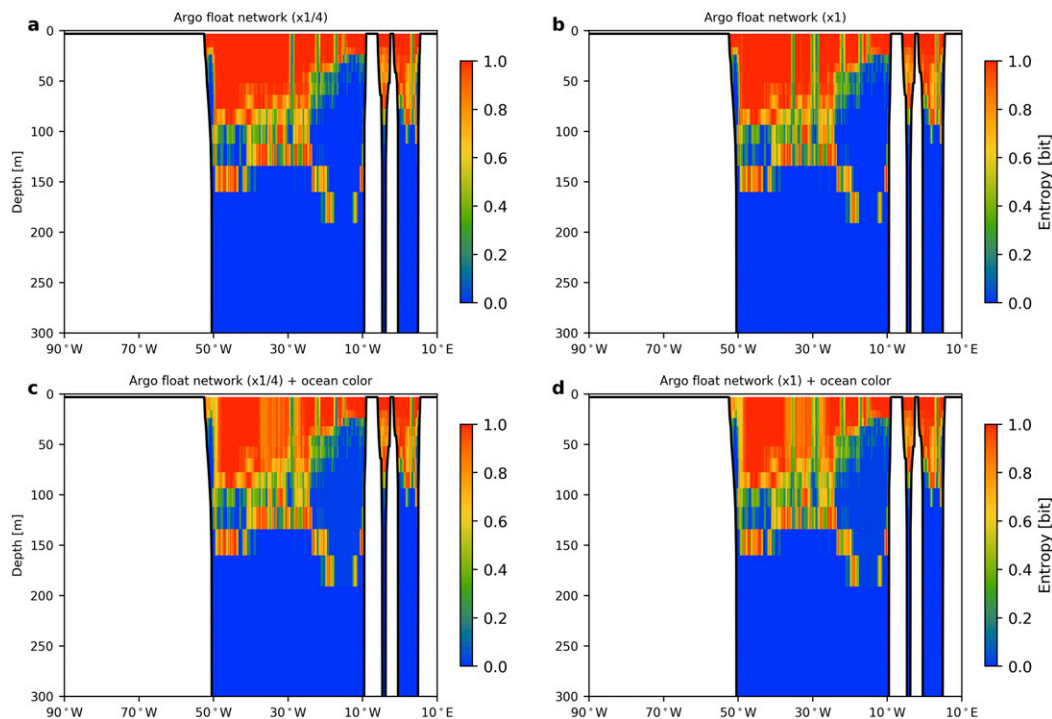


FIG. 11. Longitudinal section at 5°N of the chlorophyll IGN_n for the defined scenarios (color coded as in Fig. 9). (a) BGC-Argo sensors on a quarter of the nominal Argo array, (b) BGC-Argo sensors on the full nominal Argo array, (c) daily satellite ocean color data and BGC-Argo on 1/4 of the nominal array, and (d) daily satellite ocean color data and BGC-Argo on the nominal array.

also provides a useful framework to discriminate different observing scenarios based on information-theoretic measures such as entropy.

Nevertheless, conducting such observation impact studies will strongly depend on the characteristics of the data assimilation system (as with any related studies based on OSE/OSSE systems), and requires an ensemble of model simulations that is realistic. With our method, the reliability of the OSSEs depends on the capacity of the stochastic perturbations in the model to provide a realistic description of model errors. This may mean making substantial efforts to optimally specify the various sources of uncertainty used to produce the ensemble (i.e., in the initial conditions and in the model's equations). One must therefore keep in mind that OSSE results need to be evaluated in light of possible biases due to errors associated with the ensemble used to perform the assimilation process. The results depend also on the period defined for the evaluation of the proposed observing scenarios, as well as the type of synthetic data to be assimilated.

As part of the Horizon 2020 AtlantOS project, this new ensemble-based OSSE methodology was applied to a stochastic marine ecosystem model, in which uncertainties related to uncertain biological parameters in

the model equations have been explicitly simulated using stochastic processes. The application evaluated herein is based on a single assimilation time step of synthetic chlorophyll observations to showcase the potential of our approach in assessing the quality of future deployment scenarios of BGC-Argo arrays. An important limitation of this cross validation method is that it is rather expensive, and probably too expensive if the model is embedded in the system, as for instance if the objective is to evaluate the impact of the observation scenario on the performance of an ensemble forecast (performed after the ensemble observational update). The method would also be much more difficult to apply in a cycling experiment because of the need to control the spread of the ensemble in an assimilation context. As a first attempt to keep tractable numerical costs, one may limit the assimilation step to a subset of the ensemble size or perform multi-3D analyses to build discrete time series. As a more rigorous approach, a cost-effectiveness analysis could be considered prior to conducting the OSSEs. This analysis would allow to make trade-offs between the model computer cost due to the assimilation process (e.g., the number of select ensemble members, parameterization of the localization algorithm) and

the gain of information brought by the assimilated observations.

Acknowledgments. This study has received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement 633211 (AtlantOS). The calculations were performed using HPC resources from GENCI-IDRIS (Grant 2017-011279). Daily Ocean color product was also freely downloaded from the CMEMS database (online at <http://marine.copernicus.eu/>). This study also uses Argo float trajectories, data collected and made freely available by the International Argo Program and the national programs that contribute to it (<http://www.argo.ucsd.edu>; <http://argo.jcommops.org>). The Argo Program is part of the Global Ocean Observing System (<http://www.ioc-goos.org/>). This work benefited from stimulating and helpful comments from three anonymous reviewers to improve the manuscript. Cyril Germaineaud's work on this study was carried out in part under the auspices of the Cooperative Institute for Marine and Atmospheric Studies (CIMAS), a cooperative institute of the University of Miami and the National Oceanic and Atmospheric Administration (NOAA), Cooperative Agreement NA10OAR4320143. Cyril Germaineaud also acknowledges support from the NOAA/Atlantic Oceanographic and Meteorological Laboratory and the National Science Foundation (NSF) Grant 1537769.

APPENDIX

Acronyms

Some acronyms used in the text are listed here. Proper names (e.g., names of specific institutions, projects and systems such as ECMWF, AtlantOS, and NEMO/OPA, respectively) are not expanded in the text when first used. Note that DRAKKAR and NATL025 are not acronyms just names.

| | |
|-----------|--|
| AtlantOS | All-Atlantic Ocean Observing System |
| ECMWF | European Centre for Medium-Range Weather Forecasts |
| GODAE | Global Ocean Data Assimilation Experiment |
| NEMO/OPA | Nucleus for European Modelling of the Ocean/Océan Parallélisé |
| PISCES-v2 | Pelagic Interactions Scheme for Carbon and Ecosystem Studies, volume 2 |
| SeaWiFS | Sea-Viewing Wide Field-of-View Sensor |

REFERENCES

- Alvarez, A., and B. Mourre, 2014: Cooperation or coordination of underwater glider networks? An assessment from observing system simulation experiments in the Ligurian Sea. *J. Atmos. Oceanic Technol.*, **31**, 2268–2277, <https://doi.org/10.1175/JTECH-D-13-00214.1>.
- Arnold, C. P., and C. H. Dey, 1986: Observing-systems simulation experiments: Past, present, and future. *Bull. Amer. Meteor. Soc.*, **67**, 687–695, [https://doi.org/10.1175/1520-0477\(1986\)067<0687:OSSEPP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1986)067<0687:OSSEPP>2.0.CO;2).
- Atlas, R., 1997: Atmospheric observations and experiments to assess their usefulness. *J. Meteor. Soc. Japan*, **75**, 111–130, https://doi.org/10.2151/jmsj1965.75.1B_111.
- , E. Kalnay, W. Baker, J. Susskind, D. Reuter, and M. Halem, 1985a: Simulation studies of the impact of future observing systems on weather prediction. Preprints, *Seventh Conf. on Numerical Weather Prediction*, Montreal, QC, Canada, Amer. Meteor. Soc., 145–151.
- , —, and M. Halem, 1985b: Impact of satellite temperature sounding and wind data on numerical weather prediction. *Opt. Eng.*, **24**, 242341, <https://doi.org/10.1117/12.7973481>.
- , and Coauthors, 2015: Observing system simulation experiments (OSSEs) to evaluate the potential impact of an optical autocovariance wind lidar (OAWL) on numerical weather prediction. *J. Atmos. Oceanic Technol.*, **32**, 1593–1613, <https://doi.org/10.1175/JTECH-D-15-0038.1>.
- Aumont, O., C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen, 2015: PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.*, **8**, 2465–2513, <https://doi.org/10.5194/gmd-8-2465-2015>.
- Barnier, B., and Coauthors, 2006: Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dyn.*, **56**, 543–567, <https://doi.org/10.1007/s10236-006-0082-1>.
- Béal, D., P. Brasseur, J.-M. Brankart, Y. Ourmières, and J. Verron, 2010: Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: Implications for nonlinear estimation using Gaussian anamorphosis. *Ocean Sci.*, **6**, 247–262, <https://doi.org/10.5194/os-6-247-2010>.
- Bell, M., A. Schiller, P.-Y. Le Traon, N. Smith, E. Dombrowsky, and K. Wilmer-Becker, 2015: An introduction to GODAE OceanView. *J. Oper. Oceanogr.*, **8** (Suppl.), s2–s11, <https://doi.org/10.1080/1755876X.2015.1022041>.
- Benedetti, R., 2010: Scoring rules for forecast verification. *Mon. Wea. Rev.*, **138**, 203–211, <https://doi.org/10.1175/2009MWR2945.1>.
- Bertino, L., G. Evensen, and H. Wackernagel, 2003: Sequential data assimilation techniques in oceanography. *Int. Stat. Rev.*, **71**, 223–241, <https://doi.org/10.1111/j.1751-5823.2003.tb00194.x>.
- Bishop, C. H., 2016: The GIGG-EnKF: Ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. *Quart. J. Roy. Meteor. Soc.*, **142**, 1395–1412, <https://doi.org/10.1002/qj.2742>.
- , B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436, [https://doi.org/10.1175/1520-0493\(2001\)129<0420:ASWTET>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2).
- Brankart, J.-M., E. Cosme, C.-E. Testut, P. Brasseur, and J. Verron, 2010: Efficient adaptive error parameterizations for square root or ensemble Kalman filters: Application to the control of ocean mesoscale signals. *Mon. Wea. Rev.*, **138**, 932–950, <https://doi.org/10.1175/2009MWR3085.1>.
- , —, —, —, and —, 2011: Efficient local error parameterizations for square root or ensemble Kalman filters: Application to a basin-scale ocean turbulent flow. *Mon. Wea. Rev.*, **139**, 474–493, <https://doi.org/10.1175/2010MWR3310.1>.

- , C.-E. Testut, D. Béal, M. Doron, C. Fontana, M. Meinvielle, P. Brasseur, and J. Verron, 2012: Towards an improved description of ocean uncertainties: Effect of local anamorphic transformations on spatial correlations. *Ocean Sci.*, **8**, 121–142, <https://doi.org/10.5194/os-8-121-2012>.
- , G. Candille, F. Garnier, C. Calone, A. Melet, P.-A. Bouttier, P. Brasseur, and J. Verron, 2015: A generic approach to explicit simulation of uncertainty in the NEMO ocean model. *Geosci. Model Dev.*, **8**, 1285–1297, <https://doi.org/10.5194/gmd-8-1285-2015>.
- Bröcker, J., and L. A. Smith, 2007: Scoring probabilistic forecasts: The importance of being proper. *Wea. Forecasting*, **22**, 382–388, <https://doi.org/10.1175/WAF966.1>.
- Candille, G., C. Côté, P. L. Houtekamer, and G. Pellerin, 2007: Verification of an ensemble prediction system against observations. *Mon. Wea. Rev.*, **135**, 2688–2699, <https://doi.org/10.1175/MWR3414.1>.
- , J.-M. Brankart, and P. Brasseur, 2015: Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the North Atlantic ocean circulation. *Ocean Sci.*, **11**, 425–438, <https://doi.org/10.5194/os-11-425-2015>.
- Ciavatta, S., R. Torres, S. Saux-Picart, and J. I. Allen, 2011: Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? *J. Geophys. Res.*, **116**, C12043, <https://doi.org/10.1029/2011JC007219>.
- , R. J. W. Brewin, J. Skákala, L. Polimene, L. de Mora, Y. Artioli, and J. I. Allen, 2018: Assimilation of ocean-color plankton functional types to improve marine ecosystem simulations. *J. Geophys. Res. Oceans*, **123**, 834–854, <https://doi.org/10.1002/2017JC013490>.
- Cover, T. M., and J. A. Thomas, 2012: *Elements of Information Theory*. John Wiley and Sons, 776 pp.
- Doron, M., P. Brasseur, and J.-M. Brankart, 2011: Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical–biogeochemical model: Twin experiments. *J. Mar. Syst.*, **87**, 194–207, <https://doi.org/10.1016/j.jmarsys.2011.04.001>.
- , —, —, S. N. Losa, and A. Melet, 2013: Stochastic estimation of biogeochemical parameters from Globcolour ocean colour satellite data in a North Atlantic 3D ocean coupled physical–biogeochemical model. *J. Mar. Syst.*, **117–118**, 81–95, <https://doi.org/10.1016/j.jmarsys.2013.02.007>.
- Dowd, M., 2011: Estimating parameters for a stochastic dynamic marine ecological system. *Environmetrics*, **22**, 501–515, <https://doi.org/10.1002/env.1083>.
- Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, <https://doi.org/10.1007/s10236-003-0036-9>.
- Fontana, C., P. Brasseur, and J.-M. Brankart, 2013: Toward a multivariate reanalysis of the North Atlantic Ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data. *Ocean Sci.*, **9**, 37–56, <https://doi.org/10.5194/os-9-37-2013>.
- Ford, D., and R. Barciela, 2017: Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products. *Remote Sens. Environ.*, **203**, 40–54, <https://doi.org/10.1016/j.rse.2017.03.040>.
- Fujii, Y., K. Ogawa, G. B. Brassington, K. Ando, T. Yasuda, and T. Kuragano, 2015a: Evaluating the impacts of the tropical Pacific observing system on the ocean analysis fields in the global ocean data assimilation system for operational seasonal forecasts in JMA. *J. Oper. Oceanogr.*, **8**, 25–39, <https://doi.org/10.1080/1755876X.2015.1014640>.
- , and Coauthors, 2015b: Evaluation of the tropical Pacific observing system from the ocean data assimilation perspective. *Quart. J. Roy. Meteor. Soc.*, **141**, 2481–2496, <https://doi.org/10.1002/qj.2579>.
- Garnier, F., J.-M. Brankart, P. Brasseur, and E. Cosme, 2016: Stochastic parameterizations of biogeochemical uncertainties in a 1/4° NEMO/PISCES model for probabilistic comparisons with ocean color data. *J. Mar. Syst.*, **155** (Suppl.), 59–72, <https://doi.org/10.1016/j.jmarsys.2015.10.012>.
- Gasparin, F., and Coauthors, 2019: Requirements for an integrated in situ Atlantic Ocean observing system from coordinated observing system simulation experiments. *Front. Mar. Sci.*, **6**, 83, <https://doi.org/10.3389/fmars.2019.00083>.
- Halliwell, G. R., Jr., A. Srinivasan, V. Kourafalou, H. Yang, D. Willey, M. Le Hénaff, and R. Atlas, 2014: Rigorous evaluation of a fraternal twin ocean OSSE system for the open Gulf of Mexico. *J. Atmos. Oceanic Technol.*, **31**, 105–130, <https://doi.org/10.1175/JTECH-D-13-00011.1>.
- , V. Kourafalou, M. L. Hénaff, L. K. Shay, and R. Atlas, 2015: OSSE impact analysis of airborne ocean surveys for improving upper-ocean dynamical and thermodynamical forecasts in the Gulf of Mexico. *Prog. Oceanogr.*, **130**, 32–46, <https://doi.org/10.1016/j.pcean.2014.09.004>.
- , M. F. Mehari, M. Le Hénaff, V. H. Kourafalou, I. S. Androulidakis, H. S. Kang, and R. Atlas, 2017: North Atlantic Ocean OSSE system: Evaluation of operational ocean observing system components and supplemental seasonal observations for potentially improving tropical cyclone prediction in coupled systems. *J. Oper. Oceanogr.*, **10**, 154–175, <https://doi.org/10.1080/1755876X.2017.1322770>.
- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, [https://doi.org/10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2).
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- Hoffman, R. N., and R. Atlas, 2016: Future observing system simulation experiments. *Bull. Amer. Meteor. Soc.*, **97**, 1601–1616, <https://doi.org/10.1175/BAMS-D-15-00200.1>.
- , C. Grassotti, R. G. Isaacs, J.-F. Louis, T. Nehrkorn, and D. C. Norquist, 1990: Assessment of the impact of simulated satellite lidar wind and retrieved 183 GHz water vapor observations on a global data assimilation system. *Mon. Wea. Rev.*, **118**, 2513–2542, [https://doi.org/10.1175/1520-0493\(1990\)118<2513:AOTIOS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<2513:AOTIOS>2.0.CO;2).
- Johnson, K., and H. Claustre, 2016a: Bringing biogeochemistry into the Argo age. *Eos, Trans. Amer. Geophys. Union*, **97**, <https://doi.org/10.1029/2016eo062427>.
- , and —, Eds., 2016b: The scientific rationale, design, and implementation plan for a biogeochemical-Argo float array. Biogeochemical-Argo Planning Group Rep., 58 pp., <https://doi.org/10.13155/46601>.
- Le Hénaff, M., P. De Mey, B. Mourre, and P.-Y. Le Traon, 2008: Contribution of a wide-swath altimeter in a shelf seas assimilation system: Impact of the satellite roll errors. *J. Atmos. Oceanic Technol.*, **25**, 2133–2144, <https://doi.org/10.1175/2008JTECH0576.1>.
- , —, and P. Marsaleix, 2009: Assessment of observational networks with the representer matrix spectra method—Application to a 3D coastal model of the Bay of Biscay. *Ocean Dyn.*, **59**, 3–20, <https://doi.org/10.1007/s10236-008-0144-7>.

- Leutbecher, M., and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, <https://doi.org/10.1002/qj.3094>.
- Levitus, S., and Coauthors, 1998: *Introduction*. Vol. 1, World Ocean Database 1998, NOAA Atlas NESDIS 18, 346 pp.
- Longhurst, A., 1995: Seasonal cycles of pelagic production and consumption. *Prog. Oceanogr.*, **36**, 77–167, [https://doi.org/10.1016/0079-6611\(95\)00015-1](https://doi.org/10.1016/0079-6611(95)00015-1).
- Madec, G., 2008: NEMO ocean engine. IPSL Note du Pôle de Modélisation 27, 217 pp.
- Maritorena, S., O. H. F. d'Andon, A. Mangin, and D. A. Siegel, 2010: Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sens. Environ.*, **114**, 1791–1804, <https://doi.org/10.1016/j.rse.2010.04.002>.
- Mourre, B., P. De Mey, Y. Ménard, F. Lyard, and C. Le Provost, 2006: Relative performance of future altimeter systems and tide gauges in constraining a model of North Sea high-frequency barotropic dynamics. *Ocean Dyn.*, **56**, 473–486, <https://doi.org/10.1007/s10236-006-0081-2>.
- Oke, P., and Coauthors, 2015: Assessing the impact of observations on ocean forecasts and reanalyses: Part 2, regional applications. *J. Oper. Oceanogr.*, **8** (Suppl.), s63–s79, <https://doi.org/10.1080/1755876X.2015.1022080>.
- Peirollo, R., 2011: Information gain as a score for probabilistic forecasts. *Meteor. Appl.*, **18**, 9–17, <https://doi.org/10.1002/met.188>.
- Pham, D. T., J. Verron, and M. Christine Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.*, **16**, 323–340, [https://doi.org/10.1016/S0924-7963\(97\)00109-7](https://doi.org/10.1016/S0924-7963(97)00109-7).
- Roulston, M. S., and L. A. Smith, 2002: Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, **130**, 1653–1660, [https://doi.org/10.1175/1520-0493\(2002\)130<1653:EPFUIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1653:EPFUIT>2.0.CO;2).
- Schiller, A., and Coauthors, 2015: Synthesis of new scientific challenges for GODAE OceanView. *J. Oper. Oceanogr.*, **8** (Suppl.), s259–s271, <https://doi.org/10.1080/1755876X.2015.1049901>.
- Shannon, C. E., 1948: A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 379–423, <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Skákala, J., D. Ford, R. J. W. Brewin, R. McEwan, S. Kay, B. Taylor, L. de Mora, and S. Ciavatta, 2018: The assimilation of phytoplankton functional types for operational forecasting in the northwest European shelf. *J. Geophys. Res. Oceans*, **123**, 5230–5247, <https://doi.org/10.1029/2018JC014153>.
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677, <https://doi.org/10.1175/2555.1>.
- Stanski, H., L. Wilson, and W. Burrows, 1989: Survey of common verification in meteorology. WMO World Weather Watch Rep. 358, 114 pp.
- Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.*, **133**, 1789–1807, <https://doi.org/10.1175/MWR2898.1>.
- Toth, Z., O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Wiley, 137–163.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, <https://doi.org/10.1256/qj.04.176>.
- Visbeck, M., and Coauthors, 2015: More integrated and more sustainable Atlantic Ocean observing (AtlantOS). *CLIVAR Exchanges*, No. 67, International CLIVAR Project Office, Southampton, United Kingdom, 18–20.
- Xue, M., M. Tong, and K. K. Droegemeier, 2006: An OSSE framework based on the ensemble square root Kalman filter for evaluating the impact of data from radar networks on thunderstorm analysis and forecasting. *J. Atmos. Oceanic Technol.*, **23**, 46–66, <https://doi.org/10.1175/JTECH1835.1>.
- Xue, Y., C. Wen, X. Yang, D. Behringer, A. Kumar, G. Vecchi, A. Rosati, and R. Gudgel, 2017: Evaluation of tropical Pacific observing systems using NCEP and GFDL ocean data assimilation systems. *Climate Dyn.*, **49**, 843–868, <https://doi.org/10.1007/s00382-015-2743-6>.