

How Deep Argo Will Improve the Deep Ocean in an Ocean Reanalysis

FLORENT GASPARIN, MATHIEU HAMON, AND ELISABETH RÉMY

Mercator Océan International, Ramonville-Saint-Agne, France

PIERRE-YVES LE TRAON

Mercator Océan International, Ramonville-Saint-Agne, and Ifremer, Plouzané, France

(Manuscript received 18 March 2019, in final form 13 September 2019)

ABSTRACT

Global ocean sampling with autonomous floats going to 4000–6000 m, known as the deep Argo array, constitutes one of the next challenges for tracking climate change. The question here is how such a global deep array will impact ocean reanalyses. Based on the different behavior of four ocean reanalyses, we first identified that large uncertainty exists in current reanalyses in representing local heat and freshwater fluxes in the deep ocean (1 W m^{-2} and 10 cm yr^{-1} regionally). Additionally, temperature and salinity comparison with deep Argo observations demonstrates that reanalysis errors in the deep ocean are of the same size as, or even stronger than, the deep ocean signal. An experimental approach, using the $1/4^\circ$ GLORYS2V4 (Global Ocean Reanalysis and Simulation) system, is then presented to anticipate how the evolution of the global ocean observing system (GOOS), with the advent of deep Argo, would contribute to ocean reanalyses. Based on observing system simulation experiments (OSSE), which consist in extracting observing system datasets from a realistic simulation to be subsequently assimilated in an experimental system, this study suggests that a global deep Argo array of 1200 floats will significantly constrain the deep ocean by reducing temperature and salinity errors by around 50%. Our results also show that such a deep global array will help ocean reanalyses to reduce error in temperature changes below 2000 m, equivalent to global ocean heat fluxes from 0.15 to 0.07 W m^{-2} , and from 0.26 to 0.19 W m^{-2} for the entire water column. This work exploits the capabilities of operational systems to provide comprehensive information for the evolution of the GOOS.

1. Introduction

Since the beginning of the 2000s, the Argo international program has strongly changed our understanding of the oceanic variability by providing nearly global-scale estimates of temperature and salinity in the upper 2000 dbar (Argo Science Team 1998) and is now recognized as a key component of the global ocean observing system (GOOS). One of the most valuable contributions of the present Argo array is the

observation of climate-related ocean variability on broad spatial scales and on time scales of months and longer, making Argo a central component of the World Climate Research Program (WCRP)/Climate Variability and Predictability (CLIVAR) project. Thanks to this successful achievement, Argo is expanding its original mission by enhancing and extending observations both horizontally (e.g., in western boundary currents, along the equator) and vertically (into the deep ocean), as well as by adding biogeochemical parameters (Jayne et al. 2017).

The current deep ocean dataset is mainly composed of sparse hydrographic sections repeated every decade from oceanographic ships, such as during the World Ocean Circulation Experiment (WOCE) in the 1900s and currently Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP). Numerous studies based on these historical datasets have shown that deep patterns of mass and heat transports are key elements of the global circulation and its interactions with the

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

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-19-0208.s1>.

Corresponding author: Florent Gasparin, fgasparin@mercator-ocean.fr

DOI: 10.1175/JCLI-D-19-0208.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

atmosphere (e.g., [Rintoul 2007](#); [Purkey and Johnson 2010](#); [Talley 2013](#)). To improve our understanding of the complex coupled behavior of the climate system, it is therefore fundamental to monitor deep heat and freshwater contents globally ([von Schuckmann et al. 2016b](#)). But, while the large-scale variability of the upper ocean is today almost globally sampled, its deeper part (below 2000 m), which represents half of the ocean volume ([Wunsch and Heimbach 2014](#)), is not. The implementation of a global deep Argo array should thus provide unprecedented details of the deep ocean variability, and thus of the climate system.

In addition to observation-only datasets, ocean reanalyses are essential climatic datasets to monitor and report on past and present marine conditions ([Balmaseda et al. 2013](#)). In particular, ocean reanalyses are centerpieces of the annual Ocean State Report and of the Ocean Monitoring Indicators ([von Schuckmann et al. 2016a, 2018](#)), which have been implemented in the framework of the Copernicus Marine Environment Monitoring Service (CMEMS). To evaluate strengths and weaknesses of ocean reanalyses, several ensemble-based diagnostics have been performed to an ensemble of ocean reanalyses, as part of the Ocean Reanalysis Intercomparison Project (ORA-IP; [Balmaseda et al. 2015](#)). It has been shown that the ensemble of ocean reanalyses is consistent in representing the upper-ocean physical conditions at interannual and longer time scales, but strongly differ in representing the deeper ocean ([Palmer et al. 2017](#); [Storto et al. 2019](#); [Garry et al. 2019](#)). By comparing with historical transbasin sections, [Kouketsu et al. \(2011\)](#) showed that their assimilation system was able to reproduce observed temperature trends in some basins, but stated that the performance of the assimilation model was limited due to sparse data, illustrating the critical need of global and more frequent deep observations for ocean reanalyses.

Advances in our understanding of the climate system have thus revealed the crucial importance of deep observations from both climate research and operational centers perspectives, but quantitative assessments of the added value of such extension of the GOOS in representing climate-related processes remain limited. The main study, tackling the question of the deep Argo array design, is based on a decorrelation scale analysis from the previously mentioned historical datasets ([Johnson et al. 2015](#)) and states that 1200 floats will be necessary to represent deep ocean variability.

To provide a complementary approach from an ocean reanalysis perspective, we first identified typical uncertainty and error in ocean reanalyses. We then

performed a set of global numerical experiments to investigate how the implementation of a global deep Argo array would improve the representation of the deep ocean in ocean reanalyses, in preventing systematic model errors, and in detecting climate variability and climate trends. In line with the current implementation plan of deep Argo, toward a global array approaching 1 float per $5^\circ \times 5^\circ \times 30$ -day period ([Jayne et al. 2017](#)), this study clarifies the value of deep Argo in the context of data assimilation. Although only a limited number of studies on design requirements for the deep ocean have been published ([Kouketsu et al. 2011](#); [Chang et al. 2018](#); [Garry et al. 2019](#)), design studies are strongly required for anticipating and adapting the observation strategy to the effective gain of deep Argo for Argo users, responding to needs of climate research and operational communities.

The paper is organized as follows. Typical characteristics of temperature and salinity uncertainty/error existing in current ocean reanalyses are presented in [section 2](#). A detailed description of the experimental approach is provided in [section 3](#). The added value of a global deep Argo array for representing the mean state and the variability in the deep ocean is detailed in [section 4](#). A summary and discussion are provided in [section 5](#).

2. Temperature and salinity in ocean reanalyses

Conceptually, current ocean reanalyses can be a useful tool to investigate deep ocean variability because part of the satellite and in situ upper-ocean information is transferred to the deeper ocean through covariance procedures and physical balances on which ocean reanalyses are based. But in practice, assessing the reliability of the deep ocean representation in ocean reanalyses is difficult, because it requires a large data collection allowing us to investigate how ocean reanalyses are able to capture long-term climatic signal in deep waters. To overcome the lack of frequent and global deep ocean observations, we will determine typical characteristics of the deep ocean in ocean reanalyses based on two complementary approaches. First, an ensemble-based strategy is adopted to quantify uncertainty based on the different behavior of an ensemble of four ocean reanalyses [GLORYS2V4 (Global Ocean Reanalysis and Simulation), CMCC Global Ocean Physical Reanalysis System (C-GLORS), Fast Ocean Atmosphere Model (FOAM), and ECMWF Ocean ReAnalysis System 5 (ORAS5)] from the Global Reanalyses Ensemble Product (GREP; [section 2a](#)). Furthermore, the deployment of deep Argo floats since 2014 during pilot arrays provides a modern dataset allowing comparisons with

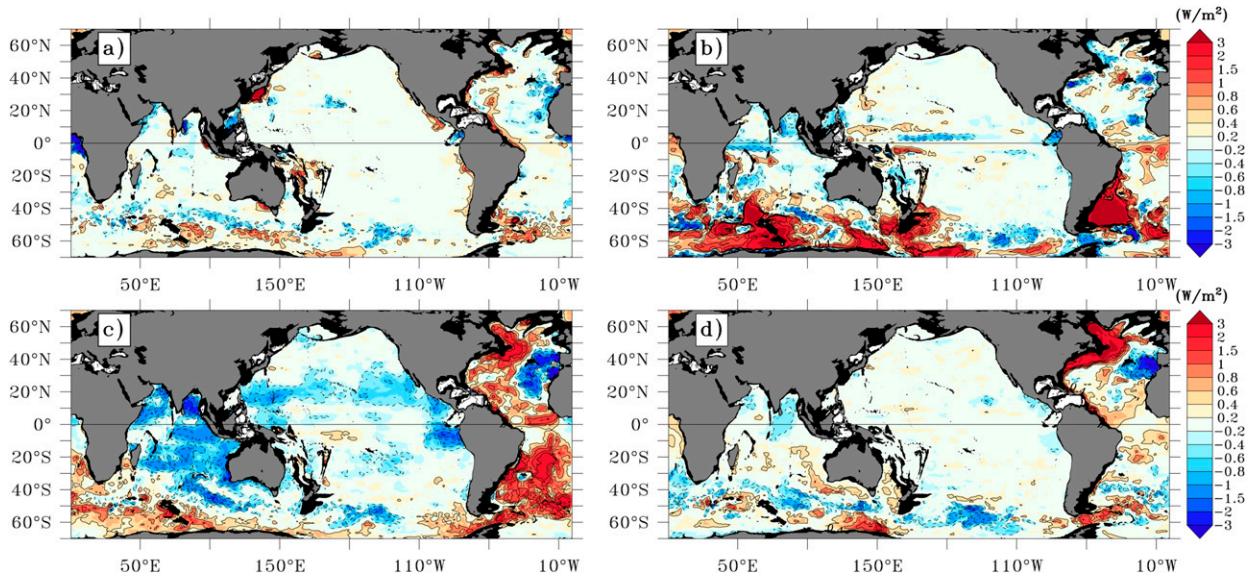


FIG. 1. Local heat content trend (W m^{-2}) within the 2000–4000-m layer over the 1993–2017 period from the GREP ocean reanalyses: (a) C-GLORS, (b) FOAM, (c) GLORYS2V4, and (d) ORAS5.

the four GREP reanalyses over the short recent period 2016–17 (section 2b).

a. Temperature/salinity uncertainty in ocean reanalyses

Covering the altimetry period, the GREP product (available in the CMEMS catalog, reference GLOBAL_REANALYSIS_PHY_001_026, data downloaded in August 2018) is composed of four global eddy-permitting ocean reanalyses using the same ocean modeling core, with a $1/4^\circ$ ORCA grid type (horizontal resolution of 27 km at the equator, 21 km at midlatitudes, and 6 km poleward), and the same atmospheric forcing dataset, ERA-Interim (hereinafter ERAi), from ECMWF (Dee et al. 2011). Differences in the observational datasets and assimilation procedures, as well as in the initialization, air–sea flux formulations, sea ice models, and process parameterizations, lead to the dispersion in the ocean-state estimates. Storto et al. (2019) detailed the main characteristics of the GREP reanalyses and demonstrated the reliability of the standard deviation over the members, defined as the spread of the ensemble, to illustrate uncertainty on the ensemble mean. Similarly, this ensemble-reanalysis system is used here to assess ocean reanalysis uncertainty on the time-mean and long-term variability over the period 1993–2017.

To illustrate the different behavior of the four ocean reanalyses, Fig. 1 shows the equivalent local heat flux into the 2000–4000-m layer from the four individual GREP members over the 1993–2017 period, following the calculation of Purkey and Johnson (2010). The

four reanalyses have important differences in the representation of the long-term heat content changes. In C-GLORS, the amplitude of temperature changes in most basins is lower than 0.2 W m^{-2} . FOAM exhibits an important warming signal especially in the Southern Hemisphere, exceeding 3 W m^{-2} regionally. In GLORYS2V4, a warming is found in the western boundary regions of the Atlantic extending to the Southern Ocean, and a cooling in the northeastern Atlantic, Pacific, and Indian Oceans. The deep ocean signal in ORAS5 is relatively weak except in the North Atlantic where patterns are similar to that of GLORYS2V4. Several reasons may explain such differences, including initialization issues (different climatological fields) and model drifts. It is noteworthy to indicate that these differences might also result from the data assimilation of the upper-ocean datasets projected onto the vertical through covariance procedures [as seen in Gasparin et al. (2018)] or from relaxation techniques to climatological fields. We refer to the work of Storto et al. (2019), who discuss why the four reanalyses might strongly differ in representing the ocean-state estimate while using the same modeling core and atmospheric datasets.

To present deep ocean changes embedded in the four ocean reanalyses of the GREP product, Figs. 2a and 2b show the ensemble mean of the four equivalent local heat and freshwater flux estimates into the 2000–4000-m layer. Positive values ($>1 \text{ W m}^{-2}$) indicate a warming around Antarctica and in the western boundary regions of the Atlantic, while negative values ($<1 \text{ W m}^{-2}$) are

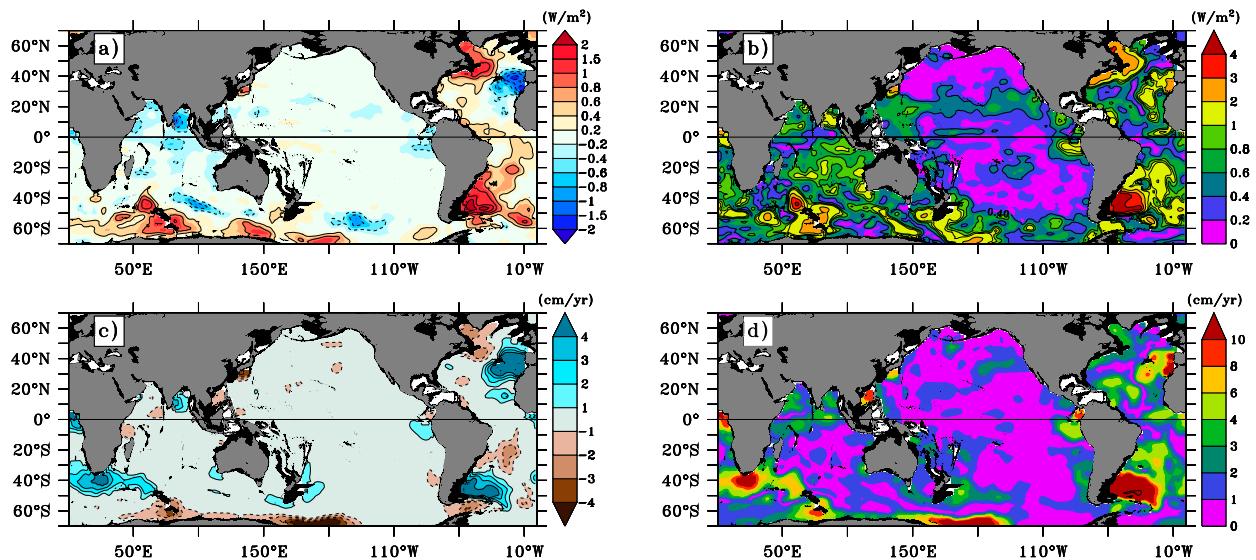


FIG. 2. (a),(c) Ensemble mean of local heat (W m^{-2}) and freshwater (cm yr^{-1}) content trends, respectively, within the 2000–4000-m layer over the 1993–2017 period from the GREP ocean reanalyses. (b),(d) As in (a) and (c), but for the ensemble spread. The ensemble mean and ensemble spread are deduced based on the trend estimates from the four reanalyses.

found in the eastern North Atlantic and to a lesser extent in the Indian and Pacific Oceans. The regional distribution of heat content changes, characterized by a strong warming in the western boundary regions of the Atlantic and in the Southern Ocean, is quite consistent with literature (Purkey and Johnson 2010; Kouketsu et al. 2011; Desbruyères et al. 2016), although regional discrepancies are found in the Indian Ocean. Similarly, the equivalent local freshwater flux, representing the amount of freshwater required for the observed dilution, is estimated by vertical integration of the 1993–2017 salinity trend over the 2000–4000-m layer multiplied by -0.03 , given that 3 cm of freshwater are needed to dilute 1 m of seawater by 1 psu, as mentioned in Gasparin and Roemmich (2016). While some waters in the Atlantic Ocean and at the south of Africa get fresher, salinity patterns suggest that the Circumpolar Deep Water and Antarctic Bottom Water get saltier in the Southern Ocean ($>2 \text{ cm yr}^{-1}$). These salinity changes have similar amplitude and opposite sign to the observed freshening of some of the Antarctic Bottom Water (Rintoul 2007; Purkey and Johnson 2013), but the confidence of the estimates from reanalyses needs to be evaluated with regard to the ensemble spread.

In Figs. 2c and 2d, the ensemble spread of local heat and freshwater fluxes, defined as the standard deviation from the four GREP estimates, exhibits a similar distribution to that of the ensemble mean (e.g., the lowest ensemble spread found in the Pacific corresponds to the lowest heat and freshwater changes in

Figs. 2a and 2b). The largest spreads are found in the other oceans and can reach more than 2 W m^{-2} and 10 cm yr^{-1} , respectively. A large spread might reflect the difficulties of representing deep ocean variability in highly variable regions without deep ocean observations (Storto et al. 2019). In general, the ensemble spread is more than 2 times higher than the amplitude of the signal (i.e., the signal-to-spread ratio is substantially below one), indicating that temperature and salinity changes are not statistically significantly different from zero. This demonstrates that, similarly to observation-only estimates (Johnson et al. 2018), deep ocean uncertainties from reanalysis estimates have similar amplitude to that of the deep ocean signal.

b. Comparison of ocean reanalyses with the deep Argo pilot arrays

As shown previously, the four ocean reanalyses have strong differences in representing deep ocean variations. Here the issue is explored further by comparing individually over the period 2016–17 the GREP reanalyses with the new generation of deep Argo floats recently deployed during deep Argo pilot arrays (Fig. 3). With around 60 active floats in 2018, the pilot arrays briefly consist in several experimental deployments, which are used for technological development (e.g., sensor calibration, uncertainty assessment, prototype test) and for scientific objectives including monitoring deep ocean warming (e.g., North Atlantic, Southern Ocean; Johnson et al. 2019). It is noteworthy

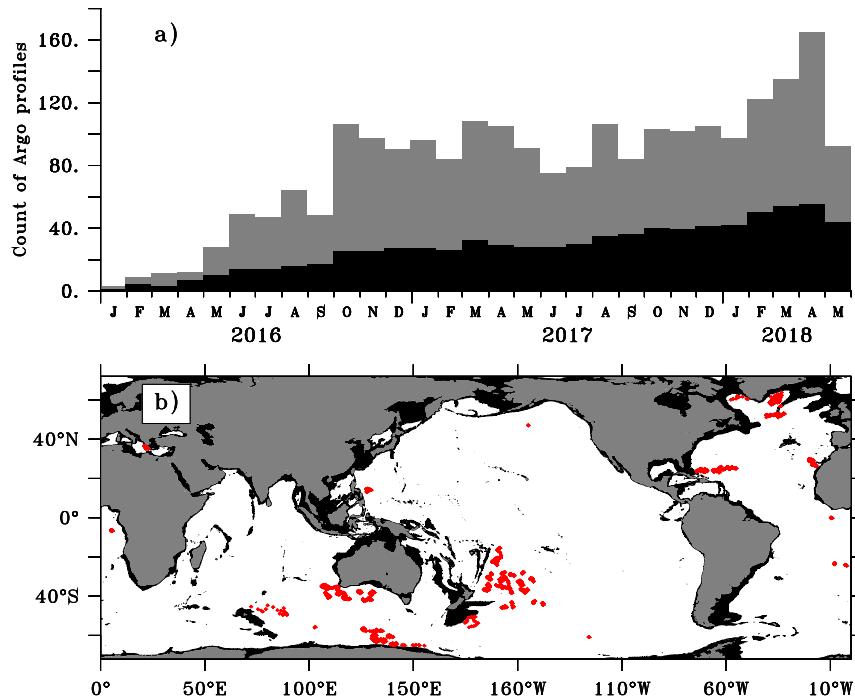


FIG. 3. (a) Count of deep Argo profiles (gray) and floats (black) from the pilot arrays for the period January 2016–May 2018. (b) Location of 1749 deep Argo profiles from the pilot arrays in 2017 (red). The deep Argo profiles shown are those that have been “adjusted” from the Global Data Assembly Center (GDAC).

that these pilot experiments are also used to define standard characteristics of the deep float sampling (i.e., cycle time, parking depth, vertical resolution) while optimizing the energy power and lifetime of the floats (Jayne et al. 2017). Because technical developments can reveal some issues in the float accuracy (e.g., pressure/salinity drifts; Le Reste et al. 2016; Roemmich et al. 2019), measurements below 2000 dbar from the deep Argo profiles are flagged by the Global Data Assembly Centers (GDAC) in such a way that they are not assimilated in GLORYS2V4 from Mercator Océan. In the C-GLORS system, vertical background-error covariances are zeroed below 2000 m, which means there is no correction at all below this depth (A. Storto 2019, personal communication). In ORAS5, deep float profiles can be assimilated but potentially due to transmission issues or specific flag determination during the pilot phase, almost no deep profiles have been assimilated in ORAS5 (H. Zuo 2019, personal communication). Unlike others, the FOAM system assimilated part of the datasets, but determining the treatment of deep Argo assimilation in the FOAM system warrants investigation that is beyond the scope of this study (M. Martin 2019, personal communication). Consequently, the analysis is based on independent

comparison using C-GLORS, GLORYS2V4, and ORAS5 in Fig. 4 while the FOAM comparison is included as supplemental material.

To be compared with reanalysis profiles, the profiles from deep floats have been converted from in situ temperature/salinity versus pressure to potential temperature/salinity versus depth. To limit the effect of spatial variability on statistics, profiles located in the North Atlantic (119 profiles) and the South Australian basin (332 profiles) are considered in Fig. 4. Including both regional and temporal variability, the standard deviation of temperature and salinity averaged in the 2000–4000-m layer from the deep Argo profiles is estimated at $7.5 \times 10^{-2} \text{C}$ and 4.2×10^{-3} in the North Atlantic and at $5.0 \times 10^{-2} \text{C}$ and 3.4×10^{-3} in the South Australian basin. The relationship between reanalysis and deep Argo estimates differs according to parameter, reanalysis, and region. In the North Atlantic, while the mean temperature difference (BIAS) and root-mean-square difference (RMSD) are smaller than the temperature variability for the three reanalyses, salinity shows higher mean and RMS differences than the variability (up to 2 times higher). In addition, the salinity BIAS and RMSD are lower for GLORYS2V4 than for C-GLORS and ORAS5 in the North Atlantic, but

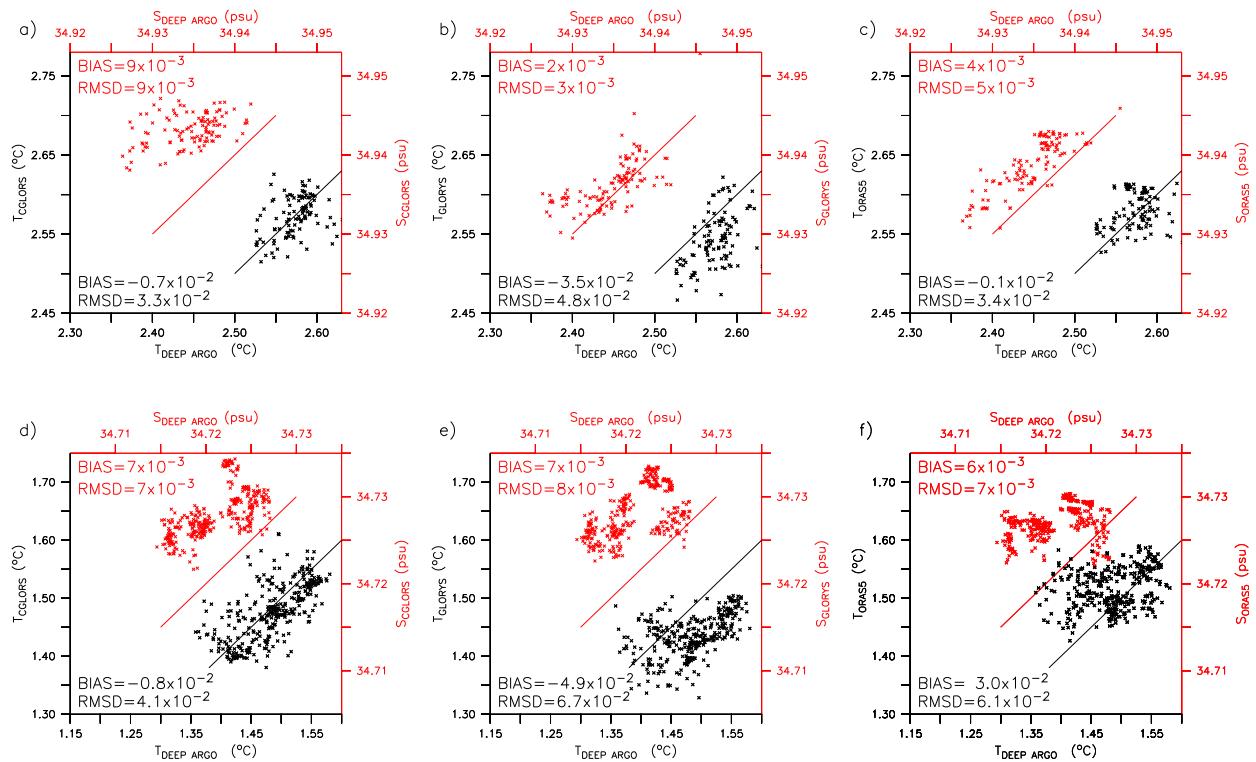


FIG. 4. Comparison of temperature (black; °C) and salinity (red), averaged over the 2000–4000-m layer, from the deep Argo profiles and the collocated GREP reanalyses profiles (C-GLORS, GLORYS2V4, ORAS5) in (a)–(c) the North Atlantic (20°–65°N, 80°–10°W; 119 profiles) and (d)–(f) the South Australian basin (30°–50°S, 100°–140°E; 332 profiles). The temperature and salinity RMS from deep Argo profiles are estimated at 7.5×10^{-2} °C and 4.2×10^{-3} in the North Atlantic and at 5.0×10^{-2} °C and 3.4×10^{-3} in the South Australian basin. See the supplemental material for comparison with the FOAM reanalysis.

not in the South Australian basin. The comparison with deep Argo datasets demonstrates that none of studied reanalyses performs better than others, suggesting that the current deep ocean observing system, mostly composed of sparse hydrographic sections repeated every decade, seems to be not sufficient to accurately monitor deep ocean changes (Desbruyères et al. 2017; Garry et al. 2019) and properly constrain ocean reanalyses at regional and global scale.

3. Numerical experimental approach

Having identified the large uncertainty and error existing in ocean reanalyses, we adopt in the rest of the paper an experimental approach in order to quantify how the implementation of a global deep Argo array would improve the ability of monitoring temperature and salinity changes in the deep ocean. A set of 6-yr numerical experiments, called observing system simulation experiments (OSSE), has been performed. OSSE consist in subsampling a “realistic” numerical simulation, called Nature Run, at the space and

time location of each observation from a given observing system design, to be subsequently assimilated into an experimental system. The ability of a given observing system to capture ocean variability is then assessed by comparing the analysis fields from the experiments with the Nature Run fields. Concretely, two main different integrated observing system designs have been subsampled from the $1/12^\circ$ Nature Run, to be assimilated in the $1/4^\circ$ experimental system. Table 1 reports the main characteristics of each simulation, including model details and assimilated datasets.

In this section, the synthetic datasets are first described (i.e., the observing system designs plus the Nature Run on which datasets have been extracted). Then, the experimental system, which corresponds to an experimental version of the GLORYS2V4 configuration, is briefly detailed. Finally, we demonstrate the consistency of our experimental approach by showing that experimental temperature and salinity errors are similar to that of current ocean reanalyses, previously described. Note that the interannual variability embedded in the Nature Run is mostly dominated by the long-term

TABLE 1. Set of simulations used in the analysis. IFS refers to the ECMWF’s operational Integrated Forecasting System fields, while ERAi refers to the ECMWF interim reanalysis fields; *WOA13* is *World Ocean Atlas 2013*. The UPPER design includes satellite (i.e., altimetry and sea surface temperature) and upper in situ components (i.e., XBT, moorings, core-Argo above 2000 m). The FULL design refers to the UPPER design plus the deep Argo component (i.e., around 1200 deep floats monthly sampling to the bottom). Note that the FULL4000 experiment is the same as the FULL experiment, except that the deep floats sample to 4000 m rather than to the bottom.

Simulation	Model characteristics				Period
	Grid resolution	Atmospheric forcing	Initialization	Assimilation	
Nature Run	1/12°	IFS	EN4	None	2007–15
UPPER	1/4°	ERAi	<i>WOA13</i>	UPPER design	2008–13
FULL	1/4°	ERAi	<i>WOA13</i>	FULL design	2008–13

trend, and diagnostics made with the 6-yr experiments could potentially be extended to longer periods.

a. Synthetic observing system datasets

Two main different integrated observing system designs have been defined, including the satellite and in situ components. The UPPER design, mimicking the current ocean observing system, is based on the same datasets than the GLORYS2V4 reanalysis (i.e., altimetry, sea surface temperature, and upper-ocean in situ datasets). No synthetic shipboard hydrographic datasets have been assimilated, but, as seen in section 2, their impact on ocean reanalyses appears to be limited. The altimetry synthetic dataset is built from a constellation of three satellites. The sea surface temperature synthetic dataset consists in daily fields released on a regular grid at 1/4° horizontal resolution. The synthetic in situ datasets rely on subsurface vertical profiles of temperature and salinity from moorings platforms, XBTs, and core-Argo floats (above 2000 dbar), which have been extracted from the CORA 4.1 in situ database (Cabanès et al. 2013; Szekely et al. 2016). The synthetic Argo component of the UPPER design has been built considering the time

and date location of existing Argo profiles during the period 2009–11, which have been combined in order to design a near-homogeneous Argo sampling, approaching 1 float per $3^\circ \times 3^\circ \times 10$ -day period.

To estimate the contribution of adding a global deep Argo array, the FULL design is composed of the UPPER design plus a deep component (equivalent to 1 float per $5^\circ \times 5^\circ \times 30$ -day period), similarly to the anticipated deep Argo design (Jayne et al. 2017). To this end, one profile per month from a third of the Argo floats of the UPPER design has been extended to the bottom (corresponding to 14921 profiles in 2010), equivalent to a near-global array of around 1200 deep Argo floats (Fig. 5). The vertical levels of the synthetic deep observations correspond to the vertical levels of the model, with 300-m resolution at 2000 m and 450-m resolution at the bottom. Note that the zonally averaged count of deep Argo floats considered in the FULL design illustrates the quite homogeneous sampling, close to 1 float per $5^\circ \times 5^\circ$ square, except south of 50°S (north of 50°N) where the existing Argo sampling is low (high) (Fig. 5b). A twin dataset of the FULL design, in which deep Argo profiles have been extended to only 4000 m, is called FULL4000 and is used in section 4a. Each

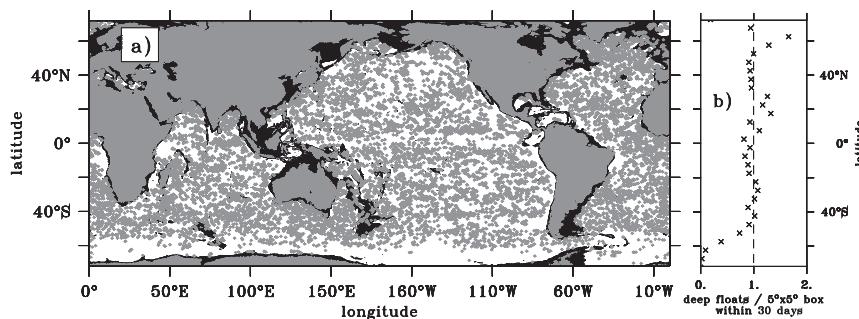


FIG. 5. Experimental FULL design (and FULL4000) for the year 2010 [(a) spatial map and (b) zonal average], corresponding to 14921 synthetic deep Argo profiles. The unit in (b) is equivalent deep Argo floats per $5^\circ \times 5^\circ \times 30$ -day period. The deep Argo profiles (~ 1110 floats) correspond to one profile per month from a third of the floats from the UPPER design, which have been extended to the bottom. The same profiles have been extended to only 4000 m in the FULL4000 design.

observation of these observing system designs has been spatially and temporally collocated to the “realistic” Nature Run daily fields, for producing synthetic datasets, in which instrumental and representation errors have been added to be consistent with the operational framework. More details of the selection profiles and added error procedures can be found in Gasparin et al. (2019).

The Nature Run (2007–15 period), in which synthetic datasets have been derived, corresponds to the unconstrained version of the global $1/12^\circ$ monitoring and forecasting system at Mercator Océan (Lellouche et al. 2018). For avoiding spurious effects due to data assimilation such as unexpected shocks or discontinuity in the analysis fields (e.g., Sivareddy et al. 2017), there is no data assimilation in this simulation. Because processes unresolved in the Nature Run cannot be evaluated in OSSE (Halliwell et al. 2014), it is fundamental to demonstrate that the deep ocean variability embedded in the Nature Run is broadly consistent with literature. A recent analysis showed that prominent features of variability are well reproduced by the Nature Run (Gasparin et al. 2018). The spatial distribution is marked by a more intense warming occurring in the Southern Ocean south of 40°S , consistently with literature based on estimations from repeat hydrographic sections (Purkey and Johnson 2010; Desbruyères et al. 2016). In addition to this regional distribution, the magnitude of the variability is also consistent with estimates from observations, which provides a good confidence of the statistical realism of the Nature Run.

b. Numerical experimental system

The UPPER and FULL (FULL4000) experiments have been performed by assimilating the UPPER and FULL (FULL4000) synthetic datasets into the experimental system. The latter corresponds to an experimental version of the GLORYS2V4 configuration and is based on the version 3.1 of the NEMO ocean model. It uses a $1/4^\circ$ ORCA grid type (horizontal resolutions of 27 km at the equator, 21 km at midlatitudes, and 6 km poleward), and has been initialized in 9 January 2008, using temperature and salinity profiles from the *World Ocean Atlas 2013* climatology (Locarnini et al. 2013; Zweng et al. 2013). To ensure that experimental errors will be similar to that of ocean reanalysis (Halliwell et al. 2014), the experimental configuration of the NEMO ocean model is different from that of the Nature Run (Madec et al. 2008). Additionally, the ocean model is forced at the surface with the atmospheric fields from ERAi produced by the ECMWF, while the Nature Run is forced by the ECMWF operational fields (IFS). Thus, the experimental system mostly differs from the Nature

Run in the horizontal resolution, atmospheric forcing, and initialization (Table 1). More details concerning parameterization of the terms included in the momentum, heat, and freshwater balances (i.e., advection, diffusion, mixing, or surface flux) can be found in Lellouche et al. (2013).

In addition to the ocean model, data assimilation procedures based on a reduced-order Kalman filter derived from a SEEK filter (SAM2; Brasseur and Verron 2006) are used for the assimilation of satellite and in situ observations. A 3D-Var bias correction for the slowly evolving large-scale error of the model in temperature and salinity is applied. More details in the data assimilation procedures can be found in Lellouche et al. (2013). Note that unlike the reanalysis, no mean dynamic topography is used for referencing the altimetric sea level anomaly since the total synthetic sea surface height is directly assimilated in the system and there is no relaxation procedure to the climatological fields. To briefly summarize, the study is mainly based on three 6-yr simulations during the period 2008–13: the Nature Run, the UPPER experiment based on the UPPER design, and the FULL experiment based on the FULL design. The contribution of an observing system to represent the deep ocean is determined through its ability to reproduce deep ocean characteristics embedded in the Nature Run. Consequently, the $1/12^\circ$ Nature Run has been interpolated to the same $1/4^\circ$ grid than experimental simulations. In the following, the differences between the UPPER experiment (the FULL experiment) and the Nature Run are referred as “UPPER errors” (“FULL errors”).

c. Consistency of the experimental approach

To validate our experimental approach, experimental errors from the UPPER experiment are compared with uncertainty and error deduced from the ensemble-based approach and the deep Argo comparison focusing on the representation of the time-mean and 5-yr changes over the period 2009–13. In Figs. 6a and 6b, the UPPER absolute temperature and salinity errors in the 2000–4000-m layer averaged over the period 2009–13 are zonally averaged and shown versus latitude (solid line). The associated temperature and salinity GREP ensemble spreads at the same level and time-averaged over the same period are also shown (dashed lines). The two curves are remarkably similar in temperature, with values of around 0.04°C north of 30°S and more than 0.1°C south of 40°S in the Southern Ocean. Similarly, there is a good consistency of the UPPER errors with the GREP spread for salinity, with values in the range 0.005–0.010, except north of 40°N where the UPPER errors are slightly higher than the GREP spread. To

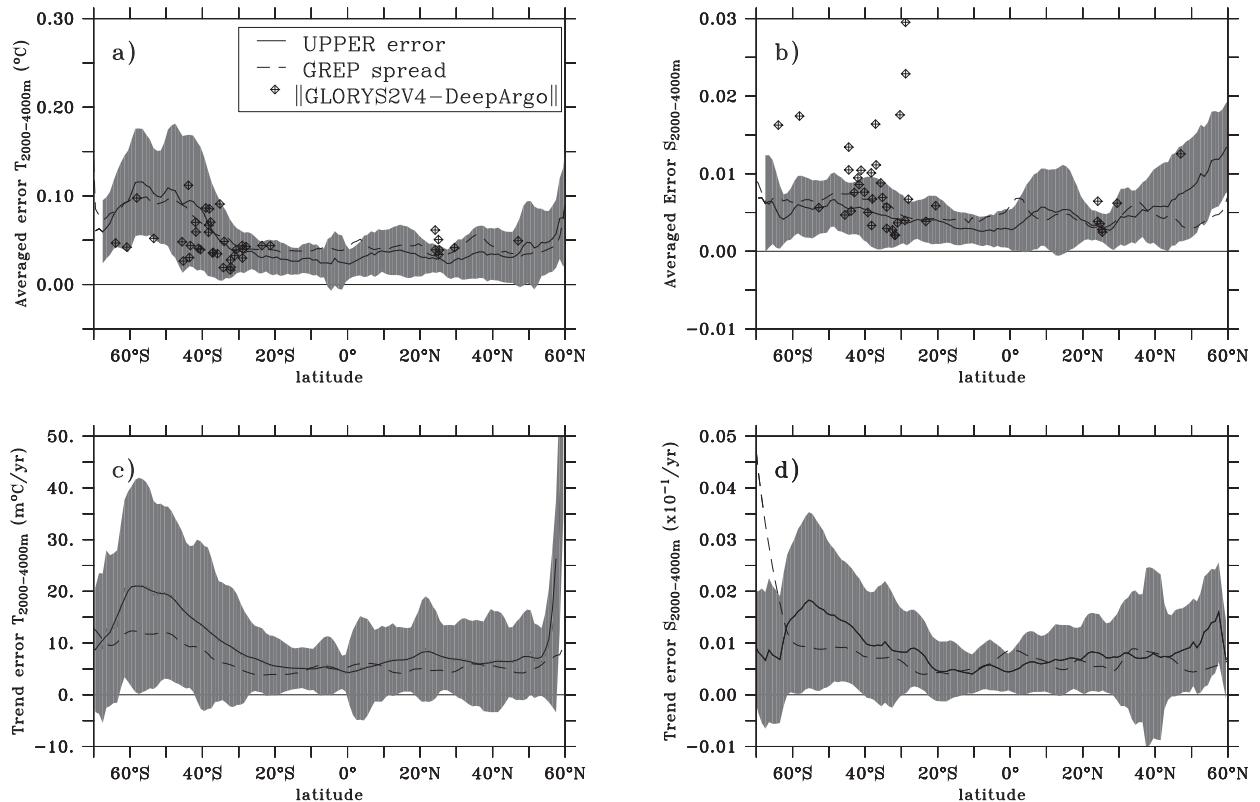


FIG. 6. (a) Zonally averaged absolute temperature error from the UPPER experiment ($^{\circ}\text{C}$) in the 2000–4000-m layer on the 2009–13 mean fields (full line) and its associated standard deviation (gray shading). The dashed line is the corresponding zonally averaged ensemble spread of the four GREP reanalyses. The squares indicate the GLORYS2V4 minus Argo differences averaged in the 2000–4000-m layer vs latitude; each square corresponds to the mean difference of modeled–observed pairs along each float trajectory. (b) As in (a), but for salinity. (c),(d) As in (a) and (b), but for the 2009–13 linear trends for temperature ($\text{m}^{\circ}\text{C yr}^{-1}$) and salinity (10^{-1}yr^{-1}).

provide other comparable estimates, the GLORYS2V4 minus DeepArgo temperature and salinity differences, averaged for each float, are plotted as a function of latitude (black squares). The 2000–4000-m temperature and salinity differences have been averaged for each float to obtain the mean difference of modeled/observed pairs for a given float. These independent error estimates have an amplitude of the same order than the UPPER errors and the ensemble spread [$O(0.1^{\circ}\text{C})$ for temperature and $O(0.01)$ for salinity], with larger variations in salinity compared to temperature. This large dispersion can reflect higher uncertainty in salinity sensors mounted on the deep floats of the pilot arrays (Le Reste et al. 2016). But, in general, most of the GLORYS2V4 minus Argo differences are included in the zonal standard deviation of the UPPER errors (gray shading).

To go further in the assessment of the consistency of our numerical approach, UPPER errors in the 2009–13 temperature and salinity changes are compared with the corresponding GREP uncertainty deduced from the ensemble spread. The UPPER error is obtained by

comparing the 2009–13 linear trend calculated from the UPPER experiment and the Nature Run, while the ensemble spread is the standard deviation of the four 2009–13 linear trends estimated from the four GREP reanalyses. As in Figs. 6a and 6b, the temperature and salinity UPPER errors at 3000 m, and the associated ensemble spread, are zonally averaged in Figs. 6c and 6d. Even if the UPPER error is higher than the ensemble spread, the two temperature and salinity error estimates have a similar latitudinal distribution marked by a stronger amplitude in the Southern Ocean (with values higher than $15 \text{m}^{\circ}\text{C yr}^{-1}$ and $15 \times 10^{-3} \text{yr}^{-1}$, respectively), and lower amplitude at the other latitudes ($5\text{--}10 \text{m}^{\circ}\text{C yr}^{-1}$ and $5\text{--}10 \times 10^{-3} \text{yr}^{-1}$). The good consistency of the UPPER errors with the GREP spread and the GLORYS2V4 minus Argo differences demonstrates that the UPPER experiment configuration is in agreement with the uncertainties to represent temperature and salinity in the deep ocean from global ocean reanalyses (e.g., Kouketsu et al. 2011), and provides good confidence in the calibration of our experimental approach and in the reliability of our sensitivity experiments

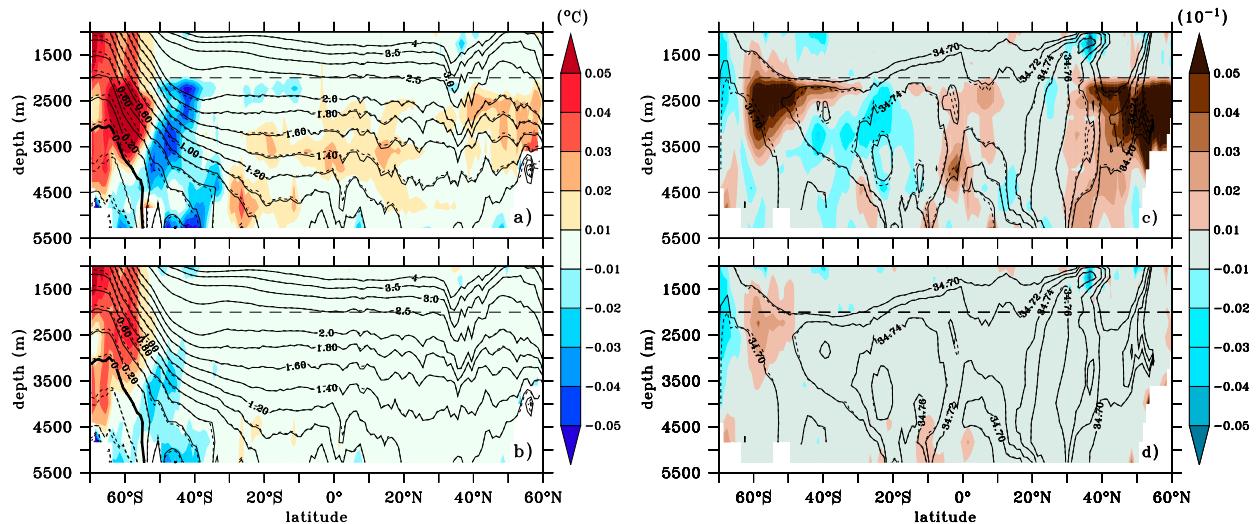


FIG. 7. (a),(b) Zonally averaged 2009–13 mean temperature error ($^{\circ}\text{C}$) vs depth from the (a) UPPER and (b) FULL experiments. (c),(d) As in (a) and (b), but for salinity ($\times 10^{-1}$). Isolines from each experiment are indicated in full black lines. The Nature Run isolines are indicated in dashed black lines.

for the deep ocean. Note that uncertainty on observation-only estimates are also of the order of several $\text{m } ^{\circ}\text{C yr}^{-1}$ (Purkey and Johnson 2010; Kouketsu et al. 2011; Garry et al. 2019; Johnson et al. 2019). Note that the same analysis carried out with the three other reanalyses give similar results.

4. Added value of a full-depth ocean observing system

To anticipate the potential contribution of deep Argo in ocean reanalyses, we first demonstrated that ocean reanalyses can struggle to appropriately estimate deep ocean variability (section 2). Then, we presented and validated a numerical experimental approach (section 3). Finally, we assess here impacts of a global deep Argo array extension using the FULL experiment, in which part of the upper-ocean observing system has been extended to the bottom, and determine how such observing system will help to capture key signals into the deep ocean.

a. Mean temperature and salinity fields

The added value of deep Argo is first assessed by evaluating its contribution to the 5-yr mean temperature and salinity fields focusing on the vertical structure and regional distribution. The monthly time series of the Nature Run and the experiments have been time-averaged over the period 2009–13. In Fig. 7, the UPPER and FULL errors have been zonally averaged, with shading indicating the UPPER and FULL error and contours showing isolines of temperature and salinity,

with dashed curves for the Nature Run and full curves for the UPPER and FULL experiments.

A net discontinuity in both temperature and salinity fields is observed at all latitudes in the UPPER experiment around 2000 m, revealing that the positive impact of the assimilation of upper-ocean datasets is mainly limited to the upper 2000 m (core-Argo's impact). The strongest temperature error is found in the Southern Ocean (having the lowest Argo sampling; Fig. 5), in the shape of a tilted dipole with warmer and colder waters to the south and to the north of 50°S , respectively. This pattern reflects a weaker latitudinal temperature gradient in the Antarctic Bottom Water in the UPPER experiment compared to the Nature Run (weaker slope of the UPPER isotherms), and thus, this should directly impact the representation of the mean geostrophic component of the Antarctic Circumpolar Current. Additionally, warmer waters are found above colder water south of 50°S , reflecting a too-strong vertical stratification in the UPPER experiment. In other regions, deep ocean waters are warmer in the UPPER experiment (deeper UPPER isotherms). The main UPPER salinity errors are seen at high latitudes (Southern Ocean, North Atlantic) and are characterized by saltier waters. In the Southern Hemisphere, saltier waters might result from a too-strong southward extension of the salinity maximum (>34.7 ; North Atlantic Deep Waters) in the formation region of the Antarctic Bottom Waters (e.g., Rintoul 2007; Yashayev 2007).

In Figs. 7c and 7d, the FULL error shows an important reduction of temperature and salinity errors compared with the UPPER errors, especially north of 40°S ($<0.01^{\circ}\text{C}$

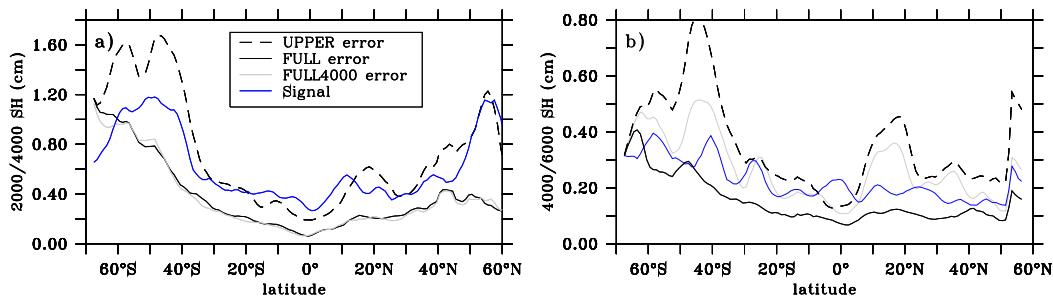


FIG. 8. (a) Zonally averaged absolute error of the mean 2000/4000 SH (cm) from the UPPER (black dashed line), FULL (black full line), and the FULL4000 (thin gray line) experiments. For comparison to the error estimate, the zonally averaged 2014 minus 2009 absolute difference of 2000/4000 SH from the Nature Run is shown in blue. (b) As in (a), but for 4000/6000 SH (cm).

and <0.001 , respectively) and to a lesser extent in the Southern Ocean ($\sim 0.03^{\circ}\text{C}$ and 0.002 , respectively). The better matching of the isolines in the FULL experiment with that of the Nature Run results from vertical adjustments of the isopycnal surfaces reaching more than 500 m. The strongest error of the thermohaline stratification is still observed south of 40°S and might reflect the lower deep Argo sampling at these latitudes as mentioned previously. Although our experiment shows that the Southern Ocean will have a stronger error compared to the other regions (likely due to the lower than targeted sampling in this region; see section 3a), this analysis suggests that the addition of deep Argo observations will improve the vertical stratification at all latitudes in replacing temperature and salinity isolines. The horizontal and vertical extensions of deep water masses should be well recovered, as seen for the North Atlantic Deep Water.

Figure 8 shows the zonally averaged absolute error of the mean steric height at 2000 m relative to 4000 m and at 4000 m relative to 6000 m (hereafter 2000/4000 SH and 4000/6000 SH, respectively) from the UPPER, FULL, and FULL4000 experiments. The FULL4000 experiment is a supplemental experiment, which has been performed to distinguish the contribution of deep Argo sampling to 4000 or to 6000 m. For comparison, the zonally averaged 2014 minus 2009 absolute difference of 2000/4000 SH from the Nature Run is shown for illustrating the amplitude of variability. In the UPPER experiment, error in the 2000/4000 SH (~ 0.4 – 0.8 cm) and 4000/6000 SH (~ 0.1 – 0.2 cm) is the same size as, or even stronger than, the variations of these quantities between 2009 and 2014. In the FULL experiment, error is significantly reduced in both 2000/4000 SH and 4000/6000 SH quantities, while it appears that the 4000/6000 SH in the FULL4000 experiment remains at the same order as that of the signal amplitude. Note that the largest errors found around 40°S and 20°N from the UPPER and

FULL4000 experiments on the 4000/6000 SH result from high errors along the western boundary of the northern tropical and southern Atlantic (not shown), as part of the deep component of the general circulation (Talley 2013), and would require further investigation in these specific regions. This brief comparison suggests that assimilated deep Argo sampling to 4000 m does not allow us to capture underneath signals in the considered reanalysis.

To assess the contribution of deep Argo on the mean circulation, the mean Atlantic meridional overturning circulation over the 2009–13 period is investigated by integrating the meridional velocity below 2000 m and from the eastern to the western boundary at four latitudes (Fig. 9). The meridional transport across these latitudes is southward varying from around 5 Sv at 50°N to 12 Sv at the three other latitudes ($1 \text{ Sv} \equiv 10^6 \text{ m}^3 \text{ s}^{-1}$), slightly lower than estimates of Ganachaud and Wunsch (2003). The meridional transport can be separated into the western boundary current and the interior pathways. An improvement is seen in the representation of the zonal variations of the meridional transport in the interior (several Sverdrups), and to a lesser extent in the western boundary current (reaching more than 20 Sv at 28°N). In general, the total southward meridional transport at each latitude is not significantly changed between UPPER and FULL experiments. The sampling might not be sufficiently dense to fully resolve the flow field in western boundary current regions (Zilberman et al. 2013), which are characterized by a small zonal extension and high variability. Another explanation would be that the AMOC in the Nature Run is not strongly different from the experimental system without data assimilation. Further experimental studies investigating the benefits of deep Argo for the meridional overturning circulation should consider these two points. The deep Argo array can however be seen as complementary to other platforms of the observing

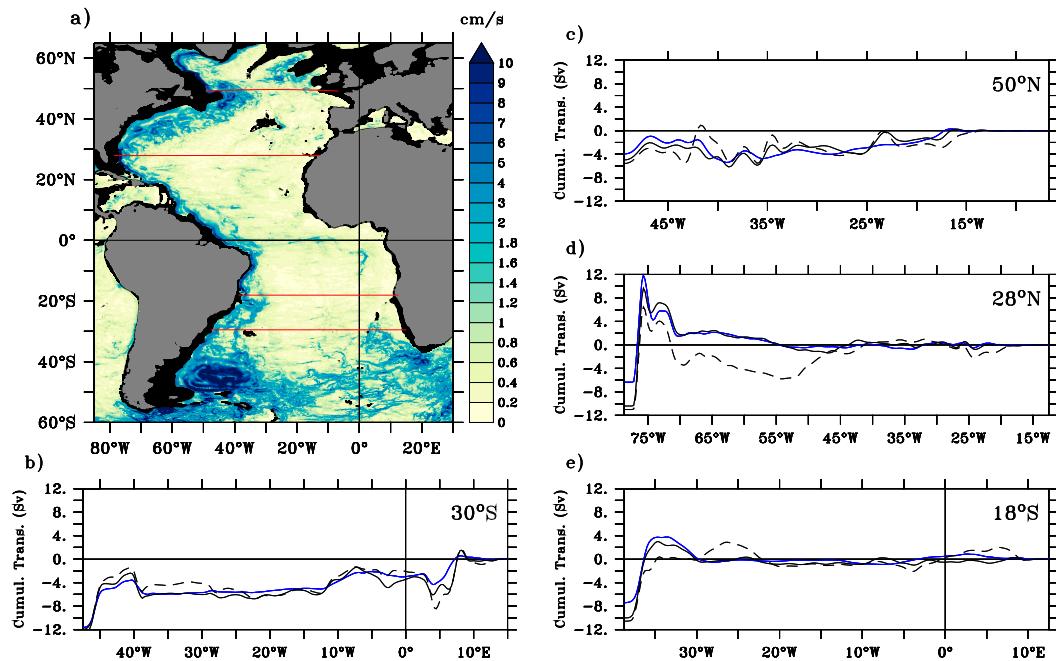


FIG. 9. Mean Atlantic meridional overturning circulation over the 2009–13 period with (a) the map of current velocity averaged over the 2000–6000-m layer (shading; cm s^{-1}), and the transbasin zonal sections (red lines), on which the 2000–6000-m meridional transport (Sv), cumulative from the eastern boundary, is calculated with (b) at 30°S, (c) 50°N, (d) 28°N, and (e) 18°S from the Nature Run (blue lines), the UPPER experiment (dashed black lines) and the FULL experiment (full black lines).

system (including moorings, gliders, and core-Argo) for monitoring the deep circulation of the AMOC (Li et al. 2017).

b. Long-term variations of deep ocean signals

Having examined the contribution of deep Argo in improving the mean temperature and salinity fields and several derived quantities, the objective here is to evaluate how deep observations will help global ocean reanalyses to capture deep ocean variability by mainly focusing on temperature changes. Similar results can be found for salinity (not shown).

By comparing the Nature Run and the experimental fields, errors in temperature, salinity, and density fields can be determined depending on monthly and annual time scales. Given the deep Argo density (i.e., $5^\circ \times 5^\circ \times 30$ -day), the two experimental and Nature Run monthly

fields were smoothed using respectively a $5^\circ \times 5^\circ \times 1$ -month and a $5^\circ \times 5^\circ \times 12$ -month running mean to represent the large-scale variability at monthly and annual time scales, respectively. Here, we use the global averaged error as the comparison metric to characterize error in temperature, salinity, and density of each experiment at monthly and annual time scales. In Table 2, the annual global averaged error of the FULL experiment for the 2000–4000-m layer (0.013°C for T , 0.012 for S , $\sim 0.011 \text{ kg m}^{-3}$ for σ), is less than half of that of the UPPER experiment ($\sim 0.031^\circ\text{C}$ for T , ~ 0.034 for S , $\sim 0.028 \text{ kg m}^{-3}$ for σ), demonstrating the positive impact of deep Argo in representing temperature, salinity, and density in the deep ocean. Unlike the UPPER error having similar amplitude at the two considered time scales, the global averaged error in the FULL experiment is smaller on longer time scales, suggesting that

TABLE 2. Globally averaged large-scale ($5^\circ \times 5^\circ$) RMS error for the UPPER and FULL experiments in temperature ($^\circ\text{C}$), salinity ($\times 10^{-1}$), and density ($\times 10^{-1} \text{ kg m}^{-3}$) for the 2000–4000-m layer at monthly and annual time scales. For the annual time scale, the monthly fields have been filtered using 12-month running means. NR refers to the Nature Run.

RMS error	Temperature		Salinity		Density	
	Monthly	Annual	Monthly	Annual	Monthly	Annual
UPPER	0.034	0.031	0.037	0.034	0.030	0.028
FULL	0.020	0.013	0.020	0.012	0.018	0.011

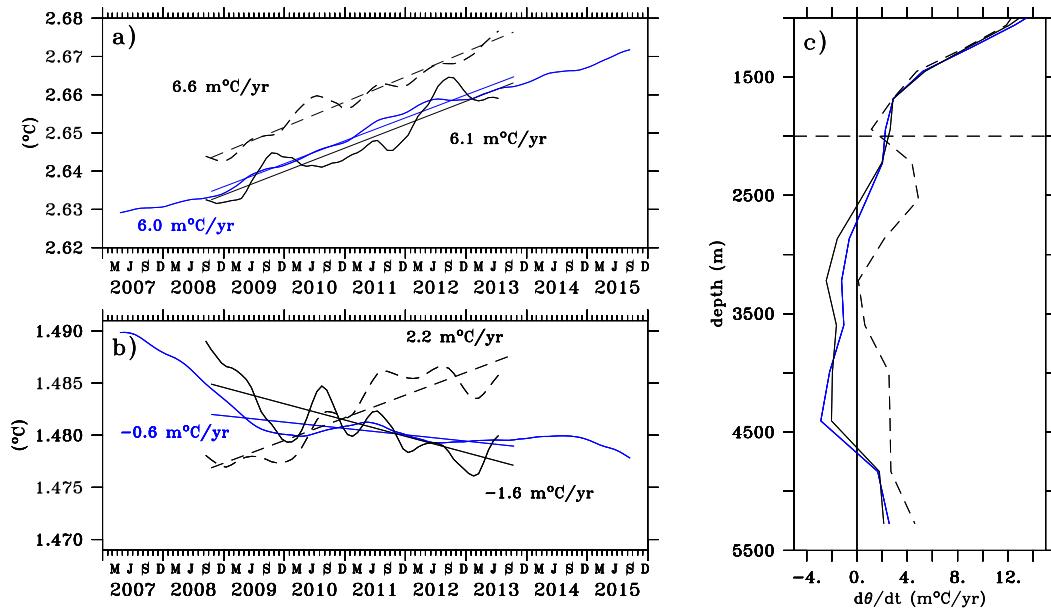


FIG. 10. (a),(b) Time series of basin-averaged 3000-m temperature in (a) the northwest Atlantic and (b) the Amundsen–Bellingshausen basins from the Nature Run (blue lines), the UPPER experiment (dashed black lines), and the FULL experiment (solid black lines) for the period 2009–13. (c) Linear trend of temperature vs depth averaged in the Amundsen–Bellingshausen basin from the same simulations. The northwest Atlantic and Amundsen–Bellingshausen basins are indicated in Fig. 11a. Note the change in the temperature range between (a) and (b).

deep Argo has a stronger impact on longer time scales. More precisely, the comparison of the typical UPPER and FULL errors suggests that the addition of deep Argo leads to an error reduction of 40% at monthly scale and around 60% at annual time scale. This is mostly related to the fact that annual variability have generally higher spatial and temporal decorrelation scales than monthly variability, implying that more Argo data will be available to capture longer scales.

To demonstrate the positive impact of deep Argo to capture long-term variability, the basin-averaged 3000-m temperature time series in the northeast Atlantic and Amundsen–Bellingshausen basins (southwest Pacific) are shown in Fig. 10 from the Nature Run, and the UPPER and FULL experiments (see Fig. 11a for the location of the two basins). In Fig. 10a, the temperature time series in the northwest Atlantic basin from the Nature Run is dominated by a 5-yr trend, estimated at $6.0\text{ m}^{\circ}\text{Cyr}^{-1}$, without shorter time scale signals. Compared to the Nature Run, temperature variations in the UPPER experiment are characterized by a bias of 0.010°C , and a similar linear trend of $6.6\text{ m}^{\circ}\text{Cyr}^{-1}$. This amplitude of the bias in the UPPER experiment is stronger than the variability over 1 year. The FULL experiment shows that the deep Argo array reduces the mean bias by a factor of 5 from 0.010° to 0.002°C , and adjusts the linear trend to $6.1\text{ m}^{\circ}\text{Cyr}^{-1}$. It is noteworthy

that annual variability is stronger in the experimental simulations, suggesting spurious effects from the vertical projection of information from the upper-ocean datasets through covariance procedures. Further investigations will be necessary to refine such procedures and adapt them to the observed variability in the deep ocean.

Unlike the northwest Atlantic basin, the UPPER time series can strongly differ from the Nature Run. As an example, the linear trend of the 3000 m-temperature time series in the Amundsen–Bellingshausen basin from the UPPER experiment ($2.2\text{ m}^{\circ}\text{Cyr}^{-1}$) is of the opposite sign compared with the Nature Run ($-0.6\text{ m}^{\circ}\text{Cyr}^{-1}$; Fig. 10b). Note that the Nature Run estimate lies in the lower limit of the Purkey and Johnson (2010) estimate (see their Fig. 7). Compared with the UPPER experiment, temperature changes are well recovered in the FULL experiment, becoming consistent with the Nature Run, with a linear trend of $-1.6\text{ m}^{\circ}\text{Cyr}^{-1}$. In Fig. 10c, the profiles of the temperature changes is shown in the Amundsen–Bellingshausen basin. It is interesting to note that the shape of the Nature Run profile, marked by a cooling between 2500 and 4500 m and a warming in the near-bottom layer, is consistent with the recent estimate of Johnson et al. (2019) based on deep Argo floats. While the UPPER experiment is characterized by an inconsistent warming below 2000 m, the FULL experiment recovered the vertical

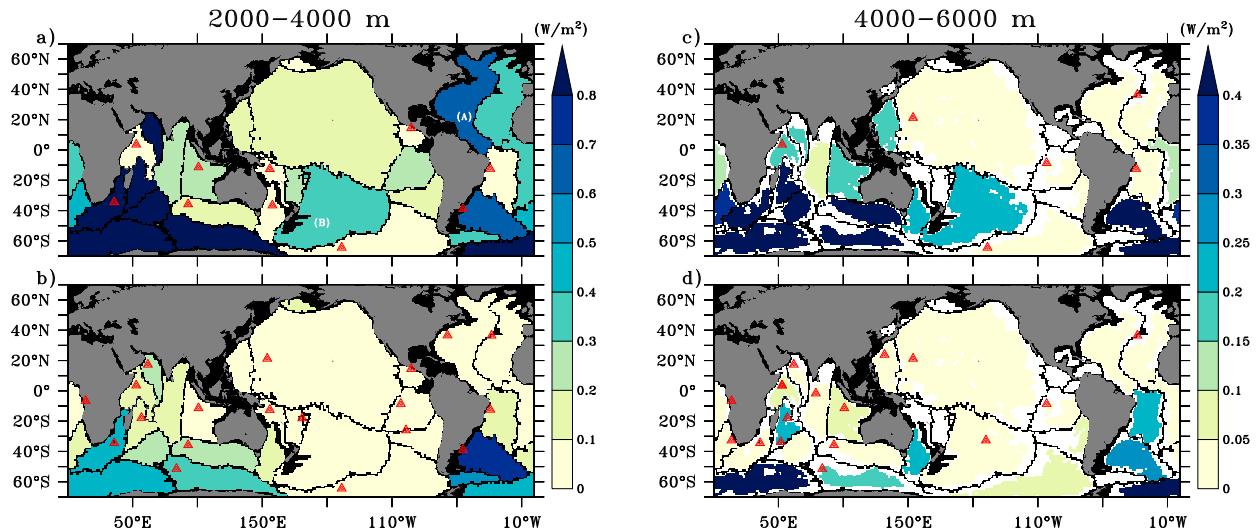


FIG. 11. Basin-averaged linear ocean heat gain (W m^{-2}) for the (a),(b) 2000–4000- and (c),(d) 4000–6000-m layers from the (a),(c) UPPER and (b),(d) FULL experiments. In (a), symbols (A) and (B) refer to the northwest Atlantic and Amundsen–Bellingshausen basins. For each basin, red triangles indicate that basin-scale signal-to-error ratio is higher than 2.

shape of temperature changes from the Nature Run. This comparison suggests that deep Argo will be able to provide statistically significant basin-scale temperature changes, improving estimates only based with the current ocean observing system (see section 2).

Figure 11 shows the differences between the 2000–4000-m heat flux from the UPPER and FULL experiments, and the Nature Run. The main discrepancies of the UPPER heat flux occur in the Atlantic and Southern Oceans, with values exceeding 0.8 and 0.4 W m^{-2} for the 2000–4000- and 4000–6000-m heat fluxes respectively. Although the regional distribution is consistent with uncertainties deduced from historical datasets (Desbruyères et al. 2017) or numerical experiment (Garry et al. 2019), the heat flux errors are slightly higher than these estimates. To interpret error with regards to the variability, basin-scale signal-to-error ratio higher than 2, defined as the ratio of the heat flux error and the Nature Run heat flux, is marked by red triangles in Fig. 11. Most of basins with signal-to-error ratio higher than 2 are located in the northern Indian Ocean. While important heat flux errors are still observed in some basins of the Atlantic and Southern Oceans, the FULL experiment induces a significant decrease of the heat flux errors in the two layers, with strong increased number of basins with high signal-to-error ratio.

Error in ocean heat and freshwater content trends is globally and basin-averaged in Table 3. The global ocean heat gain error in the 2000–6000-m layer is estimated at 0.15 W m^{-2} in the UPPER experiment decreasing to 0.07 W m^{-2} in the FULL experiment. These

errors are close to uncertainty of estimates based on observations (e.g., Llovel et al. 2014; Purkey and Johnson 2010) and suggest that error in the deep ocean heat gain could be decreased by 50% with the advent of deep Argo. In general, the deep Argo array contributes to reduce errors in basin-averaged deep ocean heat gain estimates between 45% and 80% depending on basin. As a key quantity to estimate the Earth energy imbalance (Meysignac et al. 2019), the top-to-bottom global ocean heat gain from the UPPER experiment is estimated at 0.26 W m^{-2} , which is close to uncertainty associated with estimates from in situ observations (Roemmich et al. 2015). Interestingly, the global ocean heat gain error in the FULL experiment represents a decrease of 25% of the error in the UPPER experiment.

TABLE 3. Absolute value of the mean difference in heat and freshwater fluxes for the entire water column and the 2000–6000-m layer.

Region	Simulation	HF bias (W m^{-2})		FF BIAS (mm yr^{-1})	
		0–6000 m	>2000 m	0–6000 m	>2000 m
Global	UPPER	0.26	0.15	2.1	5.6
	FULL	0.19	0.07	0.8	1.4
Pacific	UPPER	0.42	0.18	1.2	0.3
	FULL	0.26	0.03	4.3	0.3
Atlantic	UPPER	0.19	0.32	6.0	4.7
	FULL	0.16	0.13	4.5	1.7
Indian	UPPER	0.37	0.09	6.1	6.6
	FULL	0.23	0.05	8.5	4.6
Southern	UPPER	2.53	0.64	17.2	26.6
	FULL	1.76	0.12	4.1	2.9

At basin scale, deep Argo would contribute to decrease the basin-scale error in the top-to-bottom ocean heat gain between 15% and 40%. The contribution of deep Argo to freshwater gain demonstrates that at global scale the error is decrease of more than 50% for both the 0–6000- and 2000–6000-m layers. While the basin-averaged errors are systematically decreased with deep Argo in the 2000–6000-m layers, the top-to-bottom freshwater gain is not (e.g., in the Indian and Pacific Oceans). This might reflect issues in salinity representation in ocean reanalyses due to the sensitivity of ocean reanalyses to model parameterization and boundary conditions, but also to overfitting temperature observations at the expense of salinity (Storto et al. 2019).

5. Summary and discussion

Using a set of numerical experiments, the benefits of the advent a global deep Argo array for global ocean reanalyses are presented here. This numerical approach is mostly based on two experiments, called observing system simulation experiments, in which two integrated observing system datasets (i.e., including satellite and in situ components) have been extracted from a realistic simulation (Nature Run) to be subsequently assimilated in an experimental system. Both datasets include the satellite and the upper-ocean components (above 2000 m), but only one has the deep Argo extension to the bottom. The contribution of a global deep Argo array, using a sampling density similar to the anticipated deep Argo design ($5^\circ \times 5^\circ \times 30$ -day period; Jayne et al. 2017), is then assessed by evaluating the ability of the experiment based on the full-depth observing system to reproduce the deep thermohaline stratification embedded in the Nature Run, in comparison with the experiment only based on the upper-ocean observing system. Following the evolution of the GOOS, this study provides a complementary approach to the climatic research in assessing the gain of such GOOS extension from an ocean reanalysis perspective.

The first objective of the present work was to highlight the critical need of a global deep Argo array in ocean reanalyses by assessing uncertainty in ocean reanalyses in representing temperature and salinity fields in the deep ocean based on two complementary approaches, an analysis of ocean reanalysis ensemble and independent comparison with deep Argo floats. It is shown that the uncertainty is higher than the size of the signal, demonstrating that most of deep ocean changes are not statistically significant. The numerical experimental approach has then been validated using the experiment having a similar observing system to that of current

ocean reanalyses. Results demonstrate that experimental temperature and salinity errors in the deep ocean are similar to estimates determined from the ensemble-based and independent comparison approaches. This has constituted the general framework of our study, which has been used for determining the added value of implementing a global deep Argo array for ocean reanalyses.

It is shown that these new observation datasets can successfully constrain ocean reanalyses, improving the deep ocean representation of water masses. Smaller errors in temperature changes significantly increase the signal-to-error ratio. The major deep Argo achievements would be to better capture large-scale variability in temperature and salinity, and to prevent unrealistic model drifts by reducing the error in the interannual trends. For instance, due to the implementation of a deep Argo array, the model bias is reduced from 0.010° to 0.002°C in the southwest Pacific, and the interannual changes in the northwest Atlantic in the Nature Run, being in opposite sign in the absence of deep Argo, are well recovered.

Closing ocean energy and sea level budgets is critical for understanding the evolution of the climate system, and the deep ocean (below 2000 m) is known as playing an important role. Accounting for about 10% of the global full-depth ocean warming during the Argo era (Desbruyères et al. 2017), this analysis suggests that deep Argo might significantly reduce the globally and basin-averaged full-depth ocean heat gain (from 0.24 to 0.17 W m^{-2} for the global ocean), but the ability to constrain the global ocean from the surface to the bottom will raise new challenges for operational centers and dedicated investigations will be needed to fit ocean reanalyses to these new datasets.

It is important to recognize that a multisystem approach is of interest to make results more representative. Each model has its own issues, which can lead to overestimating or underestimating the contribution of an observing system in specific regions. In these regions, the sensitivity of the systems can be strongly dependent on the number of observations. More coordinated efforts of intercomparison of reanalyses such as during the ORA-IP or the AtlantOS (Optimising and Enhancing the Integrated Atlantic Ocean Observing Systems) project will improve the representativity of the results (e.g., Balmaseda et al. 2015; Oke et al. 2015; Gasparin et al. 2019). Error calculations in estimates of interannual and longer variability can be sensitive to the short period of time (Storto et al. 2017), and such studies will benefit from a longer period of time. However, our study demonstrates that the information included in deep Argo can successfully constrain ocean reanalyses.

Although there are limitations to such a numerical approach, the evolving development of reanalyses allows us to perform more sophisticated and more realistic assessments of the contributions of the observing system components compared to anterior operational systems. The increasing resolution of models might lead to a better scale-matching of ocean reanalysis with observations datasets than ever before. It is therefore essential to pursue such investigation to support the ocean observation strategy while keeping in mind the following points.

- (i) Datasets from the deep Argo pilot arrays are not systematically assimilated in global ocean reanalyses and depend on the quality control procedures applied by each operational center. Work is in progress to investigate how the assimilation of these measurements constrains numerical models, and preliminary results are consistent with the present work. Knowing that model bias might be larger than signal amplitude, questions arise concerning the assimilation of the sparsely distributed datasets during the pilot and implementation phases rather than waiting for global coverage.
- (ii) This study demonstrates that a global deep Argo array would provide a good constraint to the deep ocean in global ocean reanalyses, providing a perspective on the relevance of these observations for global ocean reanalyses. It is noteworthy that most modeling and assimilation data procedures have been developed for the upper ocean, and it will likely be necessary to refine assimilation data procedures as well as modeling development to better optimize them to the deep ocean.
- (iii) The present study suggests that the spatial density of the anticipated plan of 1 float per $5^\circ \times 5^\circ$ square will significantly benefit global ocean reanalyses in constraining the deep ocean thermohaline stratification. In the Southern Ocean, where the density sampling is the lowest, temperature and salinity errors are the strongest. Additionally, sampling to the bottom in basins deeper than 4000 m appears to provide a better estimate of the abyssal water changes than limited the sampling to 4000 m. Future scientific advances, including improvement of the representation of deep ocean processes by models, will likely allow a refinement of the effective gain of deep Argo for ocean reanalyses (e.g., in enhancing sampling in water mass formation regions).

To conclude, the present work is one of the few studies investigating the design and the relevance of global deep ocean observations. Such activities need to follow the

implementation of the global deep Argo array to take into account the scientific advances and evolving technical development of this new upcoming dataset, but it promises to provide valuable information for climate monitoring by global ocean reanalyses.

Acknowledgments. This study has been conducted using the Copernicus Marine Service Products. This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant 633211. This paper uses data collected and made freely available by programs that constitute the Global Ocean Observing System and the national programs that contribute to it (<http://www.ioc-goos.org/>), i.e., the Argo data collected and made freely available by the International Argo Project and the national programs that contribute to it (<http://argo.jcommops.org>). We thank Sarah Purkey for providing the mask of the deep ocean basins. We thank Eric Greiner and Yann Drillet and two anonymous reviewers for their helpful comments.

REFERENCES

- Argo Science Team, 1998: On the design and implementation of Argo: An initial plan for a global array of profiling floats. International CLIVAR Project Office Rep. 21, 32 pp.
- Balmaseda, M. A., K. E. Trenberth, and E. Källén, 2013: Distinctive climate signals in reanalysis of global ocean heat content. *Geophys. Res. Lett.*, **40**, 1754–1759, <https://doi.org/10.1002/grl.50382>.
- , and Coauthors, 2015: The Ocean Reanalyses Intercomparison Project (ORA-IP). *J. Oper. Oceanogr.*, **8** (Suppl. 1), s80–s97, <https://doi.org/10.1080/1755876X.2015.1022329>.
- Brasseur, P., and J. Verron, 2006: The SEEK filter method for data assimilation in oceanography: A synthesis. *Ocean Dyn.*, **56**, 650–661, <https://doi.org/10.1007/s10236-006-0080-3>.
- Cabanes, C., and Coauthors, 2013: The CORA dataset: Validation and diagnostics of in-situ ocean temperature and salinity measurements. *Ocean Sci.*, **9**, 1–18, <https://doi.org/10.5194/os-9-1-2013>.
- Chang, Y.-S., S. Zhang, A. Rosati, G. A. Vecchi, and X. Yang, 2018: An OSSE study for Deep Argo Array using the GFDL Ensemble Coupled Data Assimilation System. *Ocean Sci. J.*, **53**, 179–189, <https://doi.org/10.1007/s12601-018-0007-1>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Desbruyères, D. G., S. G. Purkey, E. L. McDonagh, G. C. Johnson, and B. A. King, 2016: Deep and abyssal ocean warming from 35 years of repeat hydrography. *Geophys. Res. Lett.*, **43**, 10 356–10 365, <https://doi.org/10.1002/2016GL070413>.
- , E. L. McDonagh, B. A. King, and V. Thierry, 2017: Global and full-depth ocean temperature trends during the early twenty-first century from Argo and repeat hydrography. *J. Climate*, **30**, 1985–1997, <https://doi.org/10.1175/JCLI-D-16-0396.1>.
- Ganachaud, A., and C. Wunsch, 2003: Large-scale ocean heat and freshwater transports during the World Ocean Circulation

- Experiment. *J. Climate*, **16**, 696–705, [https://doi.org/10.1175/1520-0442\(2003\)016<0696:LSOHAF>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0696:LSOHAF>2.0.CO;2).
- Garry, F. K., E. L. McDonagh, A. T. Blaker, C. D. Roberts, D. G. Desbruyères, E. Frajka-Williams, and B. A. King, 2019: Model derived uncertainties in deep ocean temperature trends between 1990 and 2010. *J. Geophys. Res. Oceans*, **124**, 1155–1169, <https://doi.org/10.1029/2018JC014225>.
- Gasparin, F., and D. Roemmich, 2016: The strong freshwater anomaly during the onset of the 2015/2016 El Niño. *Geophys. Res. Lett.*, **43**, 6452–6460, <https://doi.org/10.1002/2016GL069542>.
- , and Coauthors, 2018: A large-scale view of oceanic variability from 2007 to 2015 in the global high resolution monitoring and forecasting system at Mercator Océan. *J. Mar. Syst.*, **187**, 260–276, <https://doi.org/10.1016/j.jmarsys.2018.06.015>.
- , 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., 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>.
- Jayne, S. R., D. Roemmich, N. Zilberman, S. C. Riser, K. S. Johnson, G. C. Johnson, and S. R. Piotrowicz, 2017: The Argo Program: Present and future. *Oceanography*, **30**, 18–28, <https://doi.org/10.5670/oceanog.2017.213>.
- Johnson, G. C., J. M. Lyman, and S. G. Purkey, 2015: Informing Deep Argo Array design using Argo and full-depth hydrographic section data. *J. Atmos. Oceanic Technol.*, **32**, 2187–2198, <https://doi.org/10.1175/JTECH-D-15-0139.1>.
- , and Coauthors, 2018: Global oceans: Ocean heat content [in “State of the Climate in 2017”]. *Bull. Amer. Meteor. Soc.*, **94** (8), S72–S77, <https://doi.org/10.1175/2018BAMSStateoftheClimate.1>.
- , S. G. Purkey, N. V. Zilberman, and D. Roemmich, 2019: Deep Argo quantifies bottom water warming rates in the southwest Pacific Basin. *Geophys. Res. Lett.*, **46**, 2662–2669, <https://doi.org/10.1029/2018GL081685>.
- Kouketsu, S., and Coauthors, 2011: Deep ocean heat content changes estimated from observation and reanalysis product and their influence on sea level change. *J. Geophys. Res. Oceans*, **116**, C03012, <https://doi.org/10.1029/2010JC006464>.
- Lellouche, J.-M., and Coauthors, 2013: Evaluation of global monitoring and forecasting systems at Mercator Océan. *Ocean Sci.*, **9**, 57–81, <https://doi.org/10.5194/os-9-57-2013>.
- , and Coauthors, 2018: Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1/12° high resolution system. *Ocean Sci.*, **14**, 1093–1126, <https://doi.org/10.5194/os-14-1093-2018>.
- Le Reste, S., V. Dutreuil, X. André, V. Thierry, C. Renaut, P.-Y. Le Traon, and G. Maze, 2016: “Deep-Arvor”: A new profiling float to extend the Argo observations down to 4000-m depth. *J. Atmos. Oceanic Technol.*, **33**, 1039–1055, <https://doi.org/10.1175/JTECH-D-15-0214.1>.
- Li, F., M. S. Lozier, and W. E. Johns, 2017: Calculating the meridional volume, heat, and freshwater transports from an observing system in the subpolar North Atlantic: Observing system simulation experiment. *J. Atmos. Oceanic Technol.*, **34**, 1483–1500, <https://doi.org/10.1175/JTECH-D-16-0247.1>.
- Lllovet, W., J. K. Willis, F. W. Landerer, and I. Fukumori, 2014: Deep-ocean contribution to sea level and energy budget not detectable over the past decade. *Nat. Climate Change*, **4**, 1031–1035, <https://doi.org/10.1038/nclimate2387>.
- Locarnini, R. A., and Coauthors, 2013: *Temperature*. Vol. 1, *World Ocean Atlas 2013*, S. Levitus, Ed., NOAA Atlas NESDIS 73, 40 pp.
- Madec, G., and Coauthors, 2008: Nemo ocean engine. Note du Pole de Modélisation 27, Institut Pierre-Simon Laplace, 386 pp.
- Meysignac, B., and Coauthors, 2019: Measuring global ocean heat content to estimate the Earth energy imbalance. *Front. Mar. Sci.*, **6**, 432, <https://doi.org/10.3389/fmars.2019.00432>.
- Oke, P. R., and Coauthors, 2015: Assessing the impact of observations on ocean forecasts and reanalyses: Part 2, Regional applications. *J. Oper. Oceanogr.*, **8** (Suppl. 1), s63–s79, <https://doi.org/10.1080/1755876X.2015.1022080>.
- Palmer, M. D., and Coauthors, 2017: Ocean heat content variability and change in an ensemble of ocean reanalyses. *Climate Dyn.*, **49**, 909–930, <https://doi.org/10.1007/s00382-015-2801-0>.
- Purkey, S. G., and G. C. Johnson, 2010: Warming of global abyssal and deep Southern Ocean waters between the 1990s and 2000s: Contributions to global heat and sea level rise budgets. *J. Climate*, **23**, 6336–6351, <https://doi.org/10.1175/2010JCLI3682.1>.
- , and —, 2013: Antarctic Bottom Water warming and freshening: Contributions to sea level rise, ocean freshwater budgets, and global heat gain. *J. Climate*, **26**, 6105–6122, <https://doi.org/10.1175/JCLI-D-12-00834.1>.
- Rintoul, S. R., 2007: Rapid freshening of Antarctic Bottom Water formed in the Indian and Pacific Oceans. *Geophys. Res. Lett.*, **34**, L06606, <https://doi.org/10.1029/2006GL028550>.
- Roemmich, D., J. Church, J. Gilson, D. Monselesan, P. Sutton, and S. Wijffels, 2015: Unabated planetary warming and its ocean structure since 2006. *Nat. Climate Change*, **5**, 240–245, <https://doi.org/10.1038/nclimate2513>.
- , and Coauthors, 2019: Deep SOLO: A full-depth profiling float for the Argo Program. *J. Atmos. Oceanic Technol.*, **36**, 1967–1981, <https://doi.org/10.1175/JTECH-D-19-0066.1>.
- Sivareddy, S., A. Paul, T. Sluka, M. Ravichandran, and E. Kalnay, 2017: The pre-Argo ocean reanalyses may be seriously affected by the spatial coverage of moored buoys. *Sci. Rep.*, **7**, 46685, <https://doi.org/10.1038/srep46685>.
- Storto, A., and Coauthors, 2017: Steric sea level variability (1993–2010) in an ensemble of ocean reanalyses and objective analyses. *Climate Dyn.*, **49**, 709–729, <https://doi.org/10.1007/s00382-015-2554-9>.
- , and Coauthors, 2019: The added value of the multi-system spread information for ocean heat content and steric sea level investigations in the CMEMS GREP ensemble reanalysis product. *Climate Dyn.*, **53**, 287–312, <https://doi.org/10.1007/s00382-018-4585-5>.
- Szekely, T., J. Gourrion, S. Pouliquen, and G. Reverdin, 2016: CORA, Coriolis Ocean Dataset for Reanalysis. SEANO, <https://doi.org/10.17882/46219>.
- Talley, L. D., 2013: Closure of the global overturning circulation through the Indian, Pacific, and Southern Oceans: Schematics and transports. *Oceanography*, **26**, 80–97, <https://doi.org/10.5670/oceanog.2013.07>.
- von Schuckmann, K., and Coauthors, 2016a: The Copernicus Marine Environment Monitoring Service Ocean State Report. *J. Oper. Oceanogr.*, **9** (Suppl. 2), s235–s320, <https://doi.org/10.1080/1755876X.2016.1273446>.

- , and Coauthors, 2016b: An imperative to monitor Earth's energy imbalance. *Nat. Climate Change*, **6**, 138–144, <https://doi.org/10.1038/nclimate2876>.
- , P.-Y. Le Traon, N. Smith, A. Pascual, P. Brasseur, K. Fennel, and S. Djavidnia, 2018: Copernicus Marine Service Ocean State Report. *J. Oper. Oceanogr.*, **11** (Suppl. 1), S1–S142, <https://doi.org/10.1080/1755876X.2018.1489208>.
- Wunsch, C., and P. Heimbach, 2014: Bidecadal thermal changes in the abyssal ocean. *J. Phys. Oceanogr.*, **44**, 2013–2030, <https://doi.org/10.1175/JPO-D-13-096.1>.
- Yashayaev, I., 2007: Hydrographic changes in the Labrador Sea, 1960–2005. *Prog. Oceanogr.*, **73**, 242–276, <https://doi.org/10.1016/j.pocean.2007.04.015>.
- Zilberman, N. V., D. H. Roemmich, and S. T. Gille, 2013: The mean and the time variability of the shallow meridional overturning circulation in the tropical South Pacific Ocean. *J. Climate*, **26**, 4069–4087, <https://doi.org/10.1175/JCLI-D-12-00120.1>.
- Zweng, M. M., and Coauthors, 2013: *Salinity*. Vol. 2, *World Ocean Atlas 2013*, S. Levitus, Ed., NOAA Atlas NESDIS 74, 39 pp.