Project information	
Project full title	EuroSea: Improving and Integrating European Ocean Observing and Forecasting Systems for Sustainable use of the Oceans
Project acronym	**EuroSea**
Grant agreement number	862626
Project start date and duration	1 November 2019, 50 months
Project website	https://www.eurosea.eu

Deliverable information	
Deliverable number	D4.8
Deliverable title	**Assess impact of observations**
Description	Impact assessment of in situ observations for the global and North East Atlantic (IBI) Copernicus Marine modelling and data assimilation systems
Work Package number	WP4
Work Package title	Data integration, assimilation and forecasting
Lead beneficiary	Mercator Ocean International
Lead authors	Elisabeth Rémy, Romain Escudier, Alexandre Mignot
Contributors	Mercator Ocean International
Due date	30 April 2023
Submission date	28 April 2023
Comments	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862626.

https://doi.org/10.3289/eurosea_d4.8

Table of contents

Executive summary

The accuracy of the Copernicus Marine Environment and Monitoring Service (CMEMS) ocean analysis and forecasts highly depend on the availability and quality of observations to be assimilated. In situ observations are complementary to satellite observations that are restricted to the ocean surface. Higher resolution model forecasts are required by users of the CMEMS global and regional ocean analysis and forecasts. To support this with an efficient observational constrain of the model forecast via data assimilation, an increase observation coverage is needed, associated with an improved usage of the available ocean observations. This work exploits the capabilities of operational systems to provide comprehensive information for the evolution of the GOOS.

In this report, we analyse the use and the efficiency of the in-situ observations to constrain regional and global Mercator Ocean systems. Physical and biogeochemical variables are considered. The in-situ observations are used either to estimate physical ocean state at global and regional scale via data assimilation or to estimate BGC model parameters.

The impact of the physical in situ observations assimilated in open ocean and coastal areas is assessed with numerical data assimilation experiments. The experiments are conducted with the regional 1/36° resolution and global 1/12° resolution systems operated by Mercator Ocean for the Copernicus Marine Service. For the global physical ocean, the focus is on the tropical ocean to better understand how the tropical mooring observations constrain the intraseasonal to daily variability and the complementarity with satellite observations and the deep ocean.

The tropical moorings provide unique high frequency observations at different depth, but they are far away from each other, so part of the signal in the observation are decorrelated from one mooring to the others. It is only via an integrated approach, as data assimilation into a dynamical model and complementarity with other observing networks that those observations can efficiently constrain the different scales of variability of the tropical ocean circulation. As the satellite observations brings higher spatial resolution between the tropical moorings but for the ocean surface, we show that the tropical mooring and Argo profile data assimilation constrain the larger scale ocean thermohaline vertical structure (EuroSea D2.2; Gasparin et al., 2023). The representation of the high frequency signals observed at mooring location is also significantly improved in the model analysis compared to a non-assimilative simulation.

The ocean below 2000 m depth is still largely under constrained as very few observations exist. Some deep ocean basins, as the Antarctic deep ocean, shows significant trend over the past decade but they are still not accurately monitored. Based on the spread of four deep ocean reanalysis estimates, large uncertainties were estimated in representing local heat and freshwater content in the deep ocean. Additionally, temperature and salinity field comparison with deep Argo observations demonstrates that reanalysis errors in the deep ocean are of the same size as or even stronger than the observed deep ocean signal. OSSE already suggested that the deployment of a global deep Argo array will significantly constrain the deep ocean in reanalysis to be closer to the observations (Gasparin et al., 2020).

At regional and coastal scales, the physical ocean circulation is dominated by higher frequency, smaller scale processes than the open ocean which requires different observation strategy to be well monitor. The impact of assimilating high frequency and high-resolution observations provided by gliders on European shelves is

analysed with the regional Iberic Biscay and Irish (IBI) system. It was found that repetitive glider sections can efficiently help to constrain the transport of water masses flowing across those sections.

BGC ocean models are less mature than physical ocean models and some variable dependencies are still based on empirical functions. In this task, Argo BGC profile observations were used to optimize the parameters of the global CMEMS biogeochemical model, PISCES. A particle filter algorithm was chosen to optimize a 1D configuration of PISCES in the North Atlantic. The optimization of the PISCES 1D model significantly improves the model's ability to reproduce the North Atlantic bloom

Recommendations on the in-situ network extensions for real time ocean monitoring are given based on those results, and the one also obtained in the WP2, Task 2.2 where data assimilation experiments but with simulated observations where conducted. Argo extension and the complementarity with satellite altimetry was also extensively studied.

1. Introduction

In this task, we assess the impact of in situ observation in future Copernicus Marine Environment and Monitoring Service (CMEMS) ocean analysis and forecasting systems. We are considering in situ physical and BioGeoChemical (BGC) observations to be assimilated in global systems but also in a regional one since they are specific observation data sets that are important for regional ocean monitoring, especially when considering coastal regions on the shelf.
Physical in situ observations are assimilated routinely in the Copernicus Marine systems. Their impact is assessed by comparing simulation with and without assimilation, with the observations. BGC in situ observation are used to tune the model parameters.

2. Global physical observing system experiments

Data assimilation experiments are analysed to assess the role of in situ observations in constraining the global ocean. The focus is the tropical ocean and the deep ocean.

Those analysis complement the work done in WP2, task 2.2, on the complementarity of in situ and satellite altimetry to constrain the global ocean circulation at different scales. This was studied with so-called Observing System Simulation Experiment (OSSEs), an approach based on the assimilation of simulated observations from a model simulation, called Nature Run. Those OSSEs shows that the Argo temperature and salinity profiles constrain the large-scale ocean circulation and are complemented with satellite higher resolution observations that constrain the mesoscale circulation.

2.1. Tropical ocean moorings

A large spectrum of phenomena from large scale (ENSO, …) to local high frequency signal (internal tide, …) occurs in the tropical ocean. Tropical mooring observations are constraining the evolution of temperature and salinity profiles. They provided high frequency information, but they are separated by large distances. The tropical moorings are one of the most important assimilated observations in Tropical oceans, they complement the altimetric observations that constrain the dynamical topography. The model forecast should then play the role of a dynamical interpolator to give a consistent estimate of the ocean state between the mooring at spatial scales smaller than the distance between the mooring location.

To evaluate the added value of assimilating observation, an intercomparison is conducted between the CMEMS GLORYS12v1 reanalysis and the "twin" simulation but without any data assimilated. This allows to assess the impact of the assimilation and the improvement when compared to observations.

The assimilated data sets in the Glorys12 reanalysis are the CORA Delayed Time in situ data base, delivered by the Copernicus Marine in situ Thematic Assembly Center. It benefits from an improved Quality Control of the data compared to the real time observations. The assimilated observations consist of the temperature and salinity profiles from Argo floats and tropical moorings, XBT and CTD profiles from research vessels and ships of opportunity, temperature from near surface drifters and temperature and salinity time series from gliders. Temperature observations from sea mammals are also assimilated but with a higher level of error compared to the other observing networks. Only the observations with a good quality are kept in the assimilation process. Argo was shown to be the most important in situ observing network to constrain the physical ocean circulation down to 2000 m depth at global scale (V. Turpin at al., 2016). In addition, along track sea level anomalies from the Sea Level TAC (DUACS/CLS) and OSTIA SST observations are assimilated. An exhaustive description of the system can be found in Lellouche et al. 2018.

The twin reanalysis simulations, with and without assimilation, are first compared to mooring observations, and then to altimetry observations. To get an overview of the representativity of the model with and without assimilation, a spectral analysis is done for both simulations and compared to the energy in the mooring observations. This helps to better understand the scales that mooring observations can constrain in ocean analysis by data assimilation.

First, the power spectrum for the temperature is computed at the 0N-165E TAO mooring from the observations, the model simulation without assimilation (called FreeRun) and the reanalysis GLORYS12v1 (Figure 1). We focus the analysis on intraseasonal to daily frequency. Overall patterns are the same for the three estimates. The most important improvements due to data assimilation is the increase of energy at the thermocline depth, at temporal scale larger than 40 days. A slight increase can also be seen in frequency below 20 days. The buoy observations are still showing a higher level of energy at all considered time scales compared to the assimilation simulation.

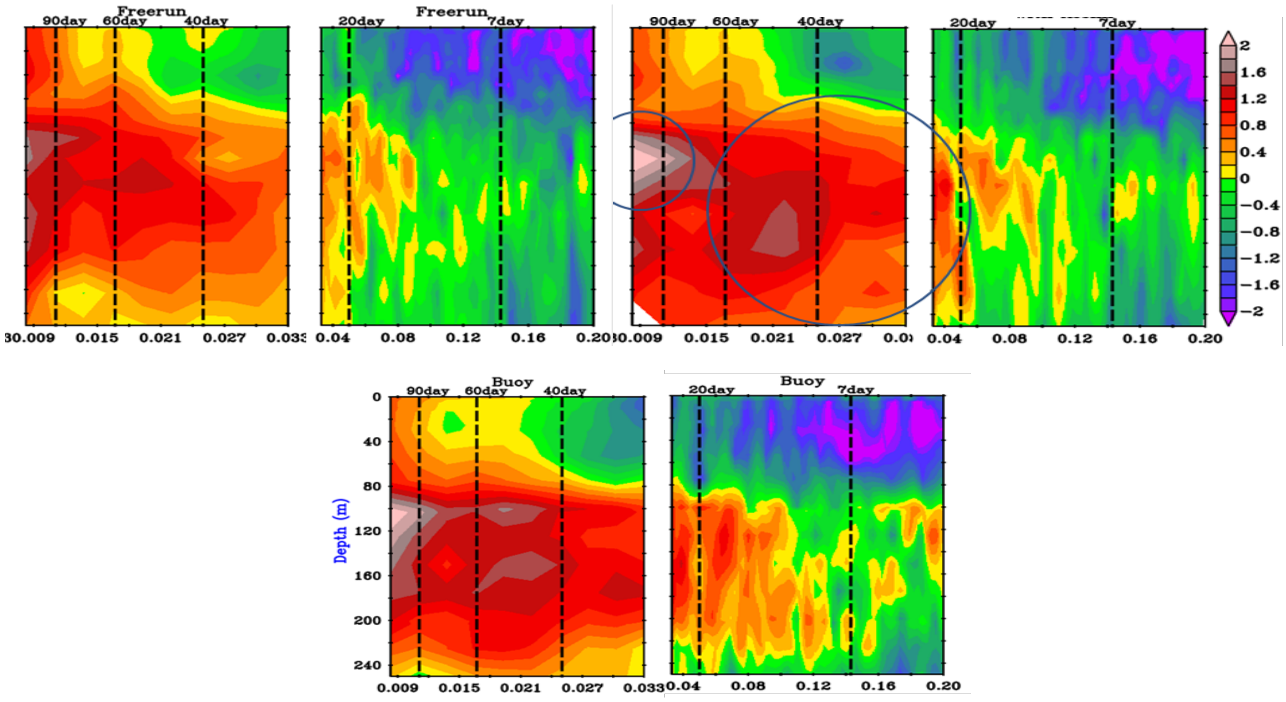

Figure 1: Spectral analysis of the temperature profile evolutions at 0N-165E: model temperature power spectrum density for the model without assimilation (top left), for the model with data assimilation (top right) and for the mooring observations. The circles highlight the depth and time frequencies the most impacted by data assimilation.

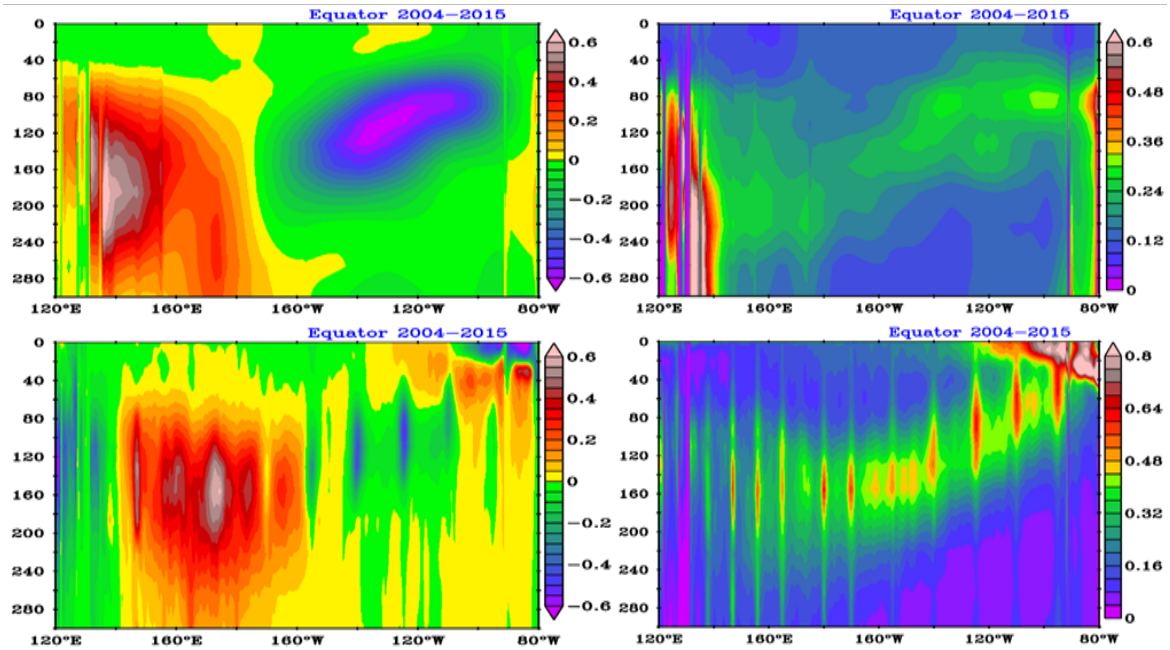

Figure 2: Equatorial vertical sections of the temperature analysis corrections over the 2004-2015 period: mean (left column) and variance (right column) of the large-scale monthly bias correction (first row) and localized weekly correction (second row) estimated by the assimilation scheme. X-axis is the longitude, Y-axis the depth in meter.

Figure 2 shows the mean and the variance of the temperature monthly large-scale analysis corrections (increments) and the localized weekly analysis corrections. The corrections are mostly localized at the thermocline depth, with opposite correction in the eastern and western part of the equatorial Pacific Ocean. The localized weekly increments show a stronger variance at the mooring location. However, when looking at the variance of the analysis fields, those corrections are dumped and do not lead to spurious circulation patterns.

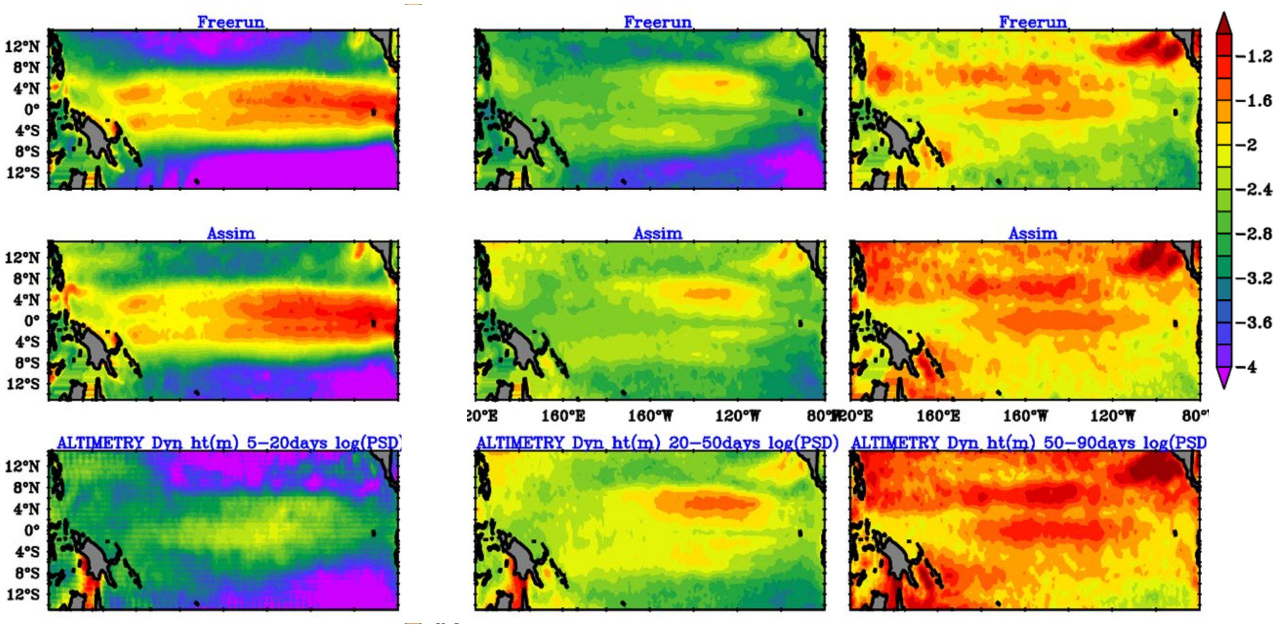

Figure 3: Dynamic Height (m) Power Spectrum [log (PSD)] for 5-30, 30-60 and 60-90 days

Figure 3 shows maps of the mean power spectrum for different frequency band, for altimetry maps (DUACS L4 SLA) and the dynamic height of the free and assimilated simulations. This intercomparison relies on the high correlation at the equator between the variability of the Dynamic Height (Steric Height) et the Sea Level Anomaly. Within the frequency range from 5 to 20 days, the assimilation slightly increases the level of energy. The altimetry maps have a lower energy due to the way they are produced by collecting observations over 10-day period. For the 20-to-50-day period, a slight increase due to the assimilation is seen, but the level of energy is still lower than from altimetry maps. The maximum energy is situated where the Tropical Instability Waves (TIWs) are the dominant process in this frequency band. For frequency between 50 to 90 days, again the assimilation increases the level of energy, to be closer to the altimetry maps.

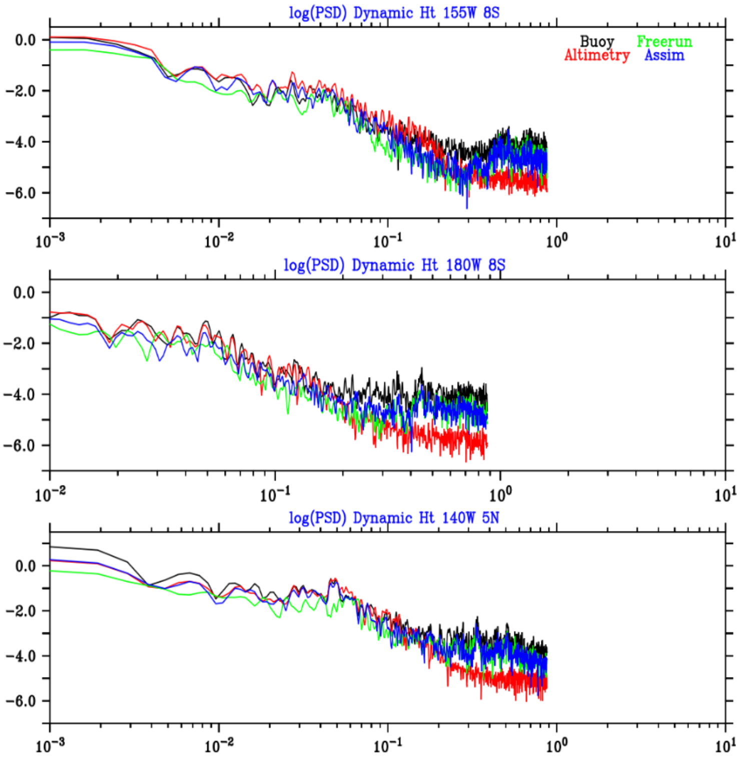

Figure 4: Intercomparison of Power Spectrum density of the Dynamic Height estimates at different mooring locations: in black, from the moorings, in blue from the assimilated simulation, in green the simulation without assimilation, in red from the DUACS altimetry maps.

The energy spectrum at three mooring locations is estimated from mooring observations and the altimetry observations, as well as the model simulations (Figure 4). This allows to verify the coherency between all estimates from in situ and altimetry observations and model simulations with and without data assimilation. Below 10 days, the energy from altimetry maps is the lower, as expected. For longer periods, buoys and assimilation simulations agree very well. The simulation without data assimilation (FreeRun) often shows a lower level of energy. This analysis agrees with the previous results.

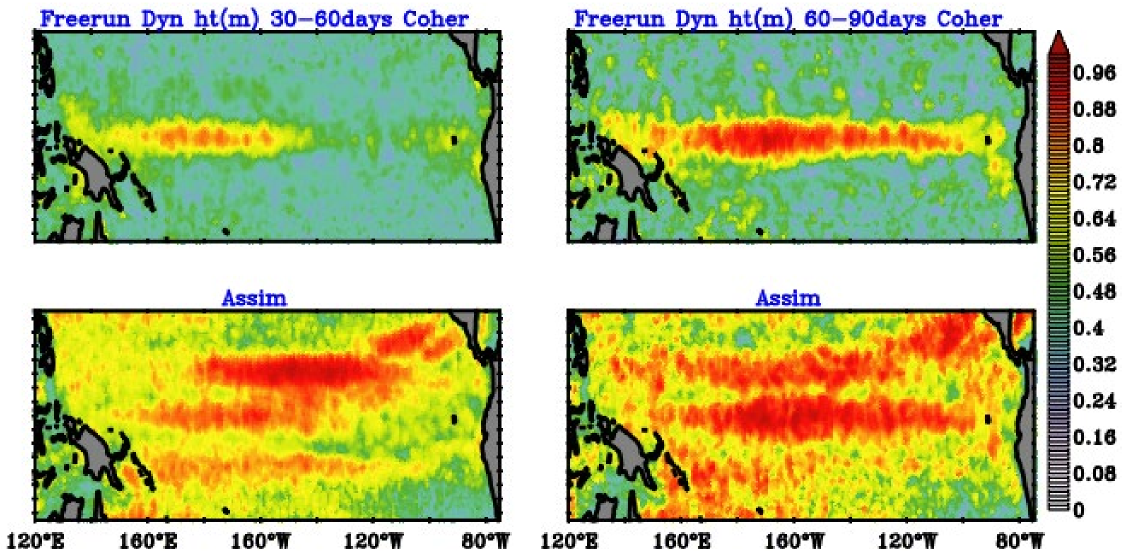

Figure 5: Coherence of Dynamic Height (0-2000m) in FreeRun and Assimilation run with respect to AVISO Altimetry Absolute Dynamic Topography maps (m) variability

The coherence is computed between the two model simulations and the DUACS absolute dynamic topography. The coherence corresponds to the correlation between time series, at different scales. Figure 5 shows the map of the coherence between the altimetry observed maps and the model estimates of the steric height with and without assimilation, for the different frequency bands. This diagnostic clearly highlights that even if the level of energy was satisfying in the free simulation, the temporal correlation of the variability was very low: the variability in the simulation without assimilation is not in phase with the observations. The model without assimilation shows a good coherency with the observed dynamic height only along the equator. The assimilation strongly improves the coherency, mostly in the region dominated by the TIW for 30 to 60 days frequency, and at the equator and (5°N) for frequency between 60 to 90 days.

Those analysis shows that assimilation of in situ and satellite data into the global 1/12° system increases the level of energy from daily to intraseasonal scales, especially at the thermocline depth. The data assimilation is efficiently constraining the reanalysis variability to be in phase with the observed one, which is not the case for the simulation without assimilation.

The in-situ observations are key to constrain the thermohaline vertical mean and variability as altimetry helps to fill the gap in space between mooring locations and inform on the steric height variability of the full water column. The numerical ocean model acts as a dynamical interpolator between the high frequency profiles observed at tropical moorings that are quite far away from each other.

2.2. Deep ocean circulation

The deep ocean is largely under constrained in ocean reanalysis, and the accuracy of the deep ocean estimate in reanalysis is not often evaluated. The deep Argo floats are still few and localized in the regions of deployment as the research vessel deep CTD lines have a very low time repeatability in time that prevent them to efficiently constrain the time evolution of the deep ocean over few decades in reanalysis. Ocean monitoring and forecasting systems usually have assimilation time window between a week and a day and the information brings by the deep CTD lines is not kept on long time scales. In order to estimate the uncertainty associated to the deep ocean freshwater and heat budget, an ensemble of 4 Copernicus global ocean reanalysis was intercompared and evaluated against deep Argo observations. As the sparsity of ocean observations deeper than 2000 m depth prevent to get a reliable global large-scale uncertainty estimate, the spread of the ensemble of reanalysis is used as a proxy for model uncertainties. All of those reanalyses are at ¼° resolution but differs in their assimilation scheme. Further description of those reanalyses is publicly available online[1]. Additionally, the GLORYS12 global 1/12° reanalysis is considered. Figure 6 shows the 3500 m depth potential temperature estimation for 4 reanalyses compared to deep Argo observations. All reanalyses show the same pattern with colder temperature along the topography compared to the open ocean but the values can differ from more than 0.2°C. It can also be noticed than meso scale patterns exist at that depth.

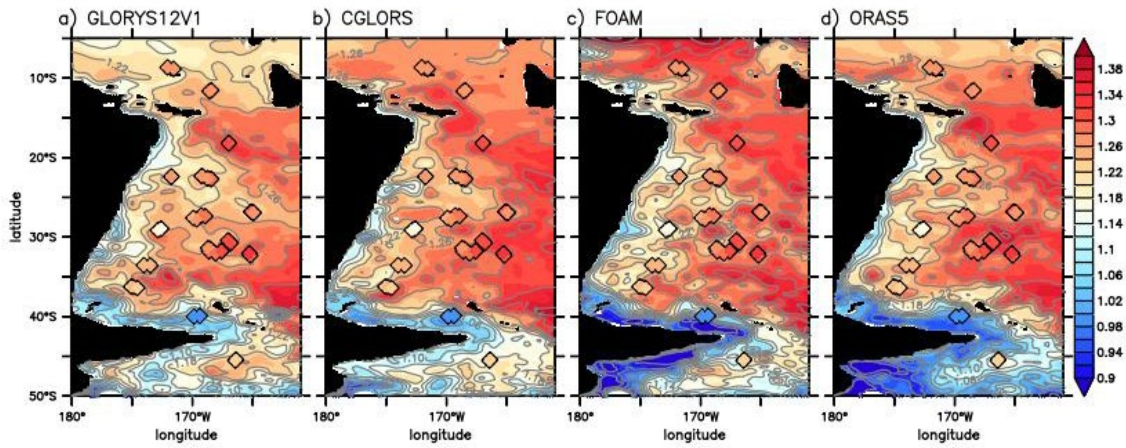

Figure 6: 3500-m potential temperature in June 2019 from (maps) different Reanalysis and (square) the deep Argo pilot array in the southwest Pacific. Only qualified deep Argo data have been considered (adjusted).

The time evolution of the temperature at 5000 m depth estimated from the different CMEMS reanalysis at ¼° resolution and the deep Argo observations is shown in Figure 7. Very different temperature trends, even with an opposite sign, are found in the different reanalysis and from the deep Argo observations.

This analysis shows the critical need to have deep regular Argo observations, not only for assimilation but also for operational product and model development evaluation. It also highlights the need of refining the assimilation strategy in the deep ocean where the time and spatial scales of the variability are much larger than in the mixed layer. Having more frequent and denser deep observations will also allow to provide reliable uncertainty estimate on the operational products distributed in the context of the Copernicus Marine service.

[1] https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-GLO-QUID-001-031.pdf

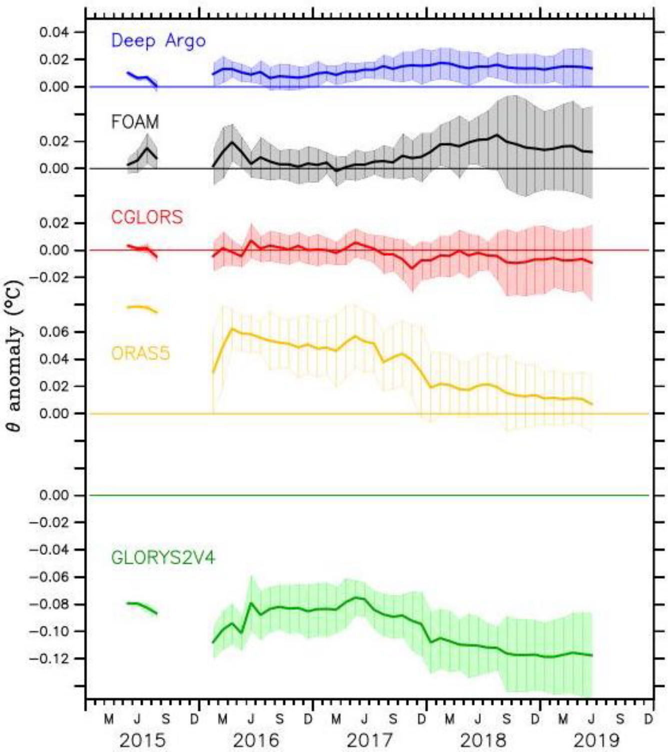

Figure 7: Different time-evolution of the 5000-m θ anomaly in Copernicus Marine GREP reanalysis (anomaly referenced to WOA18)

This intercomparison between reanalysis and deep Argo observations validate the design and complement the data assimilation experiments done with simulated observations during the AtlantOS H2020 project. The AtlantOS data assimilation experiments were produced to assess the potential of the planned deep Argo extension to improve ocean analysis. They are based on assimilation of simulated observations extracted from a model simulation called Nature Run, this approach is called OSSE for Observing System Simulation Experiment. The contribution of a global deep Argo array, using a sampling density like the anticipated deep Argo design (5° x 5° x 30-day period), was 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. Figure 8 shows the difference between the Nature Run, which should represent "the true ocean", and the deep mean temperature in the experiments without (top) and with (bottom) deep Argo observation assimilated. Those differences have the same order of magnitude (0.01 – 0.05°C) than the differences found in this task between the real observations and the model simulations, which validates the design of the AtlantOS OSSE experiments. The OSSEs showed that the deep Argo planned network can successfully constrain ocean reanalyses, improving the deep ocean representation of water masses. Smaller analysis errors in temperature changes will significantly improve the signal-to-error ratio for the climate trend estimation. 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 (Gasparin et al., 2020).

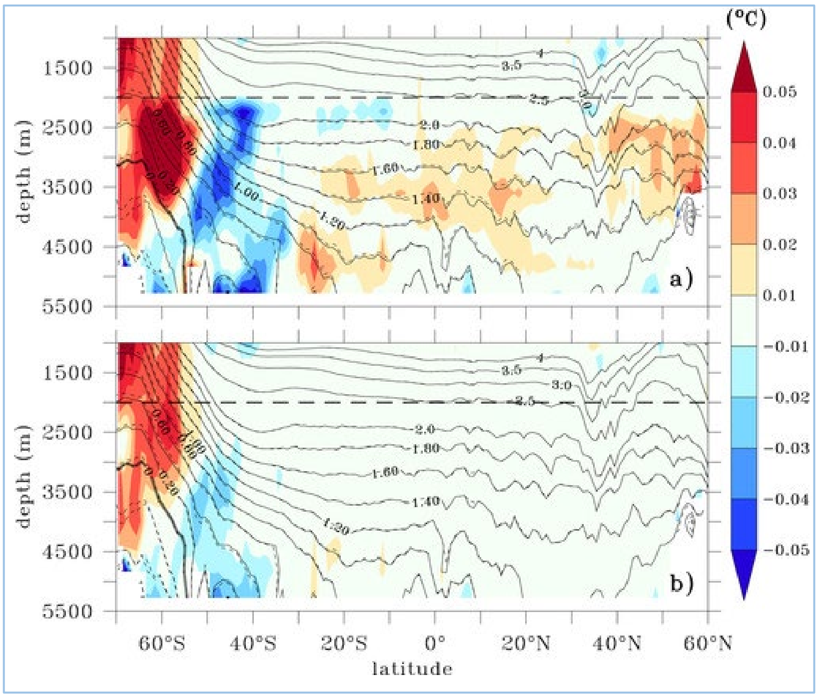

Figure 8: Zonally averaged 2009–13 mean temperature error without (top) / with (bottom) deep Argo float planned array (5° x 5° x 30-day) assimilated in the global ¼° system

2.3. Conclusion

The tropical moorings provide unique high frequency observations at different depth, but they are far away from each other, so part of the signal in the observation are decorrelated from one mooring to the others. It is only via an integrated approach, as data assimilation into a dynamical model and complementarity with other observing networks that those observations can efficiently constrain the tropical ocean circulation. The spectral analysis of the GLORYS12 1/12° global reanalysis clearly shows a great improvement of the representativity of the tropical dynamic, dominated by propagating waves, when observations are assimilated compared to a simulation without assimilation. Not only the variability but also the phase of the tropical waves is improved, this is the case from intra seasonal to daily time scales. As the satellite observations brings higher spatial resolution between the tropical moorings but for the ocean surface, the tropical mooring and Argo profiles constrain the larger scale ocean thermohaline vertical structure (EuroSea D2.2; Gasparin et al., 2023) which influence the propagation speed for the tropical waves.

The ocean below 2000 m depth is still largely under constrained as very few observations exist. Some deep ocean basins, as the Antarctic deep ocean, shows significant trend over the past decade but they are still not accurately monitored. Four CMEMS reanalysis based on different data assimilation schemes and model configuration set up were considered to assess the uncertainty of the deep ocean temperature and salinity fields. An intercomparison was conducted against each of the four reanalysis and deep Argo observations. Based on the spread of those four deep ocean reanalysis estimates, large uncertainties were estimated in representing local heat and freshwater content in the deep ocean. Additionally, temperature and salinity field comparison with deep Argo observations demonstrates that reanalysis errors in the deep ocean are of the same size as or even stronger than the observed deep ocean signal. The sparsity of deep ocean observations prevents to have reliable estimate of the deep ocean mean and variability. OSSE suggested that the deployment of a global deep Argo array will significantly constrain the deep ocean in reanalysis to be closer to the observations (Gasparin et al., 2020).

3. Regional physical observing system experiments

The Observing System Experiments (OSE) for the global system are focusing on the impact of observing networks relevant to monitor the variability of the open ocean at global scale, here we focus on a regional CMEMS ocean monitoring and forecasting system to assess and improve the benefit of assimilating observations on the shelf. Compared to the open ocean, the variability on the shelf is at higher resolution and frequency and, in parallel, observing platforms are quite different, with different sampling strategy. Different processes are driving the coastal ocean variability, such as tide, compared to the open ocean. In this task, we are trying to better understand and improve the assimilation of coastal observations, namely gliders data, in the regional Iberian Biscay Irish (IBI) system, which also covers a part of the western Mediterranean Sea (Figure 9), to better understand and increase their benefit in regional analysis.

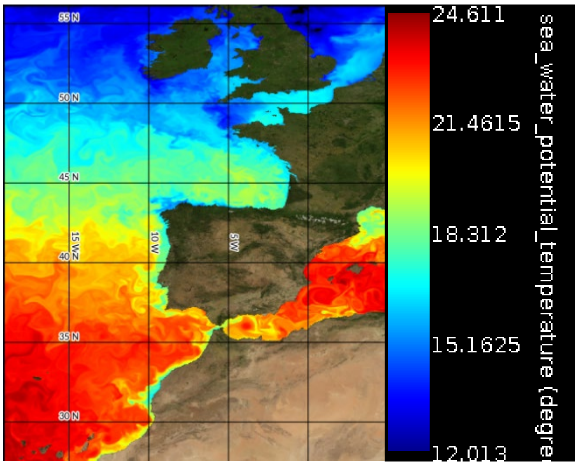

Figure 9: IBI domain, with a daily mean temperature field.

The operational IBI Ocean Analysis and Forecasting system (daily run by Nologin in coordination with Puertos del Estado and with the support, in terms of supercomputing resources, of CESGA) provides a 5-day hydrodynamic forecast including high frequency processes of paramount importance to characterize regional scale marine processes (i.e. tidal forcing, surges and high frequency atmospheric forcing, freshwater river discharge, wave current coupling in forecast, etc.). A weekly update of IBI downscaled analysis is also delivered as historic IBI best estimates. The system is based on a (eddy-resolving) NEMO model application run at 1/36º horizontal resolution, with 50 vertical levels. The data assimilation system (Mercator Ocean assimilation system SAM2) allows constraining the model in a multivariate way toward satellite Sea Surface Temperature (SST), together with all available satellite Sea Level Anomalies (SLA), and in-situ observations profiles and times series of temperature and salinity.

3.1. Experiment design

For the study, the selected period is the year 2017. In this year, several glider missions were available in the Western Mediterranean and we chose to focus in this area in order to be able to compare with the experiments performed in task 4.2. We decided to run 1 year of 3 different simulations, one free-run (FREE-R) to see how the model perform without data assimilation, one control run (NOGLID) with the data assimilation of the operational model and one "glider" run (GLIDER) adding the assimilation of gliders observations (see Table 1).

Table 1. Assimilated observations for each run.

	FREE-R	NOGLID	GLIDER
SLA	no	yes	yes
T/S from in-situ	no	yes	yes
Gliders	no	no	yes

The dataset used for the assimilation of gliders was provided by SOCIB within the framework of task 4.2 in order to be able to do the inter-model comparison later. Only the ascending profiles were kept for the assimilation and the data is interpolated on the model levels as for the other profiles that are assimilated by the system.

The assimilation scheme was not changed to integrate the new observations as we wanted to see how the system could ingest this data as it is. Only the error of the observations was changed to reflect the higher density of these profiles.

3.2. Results

General statistics

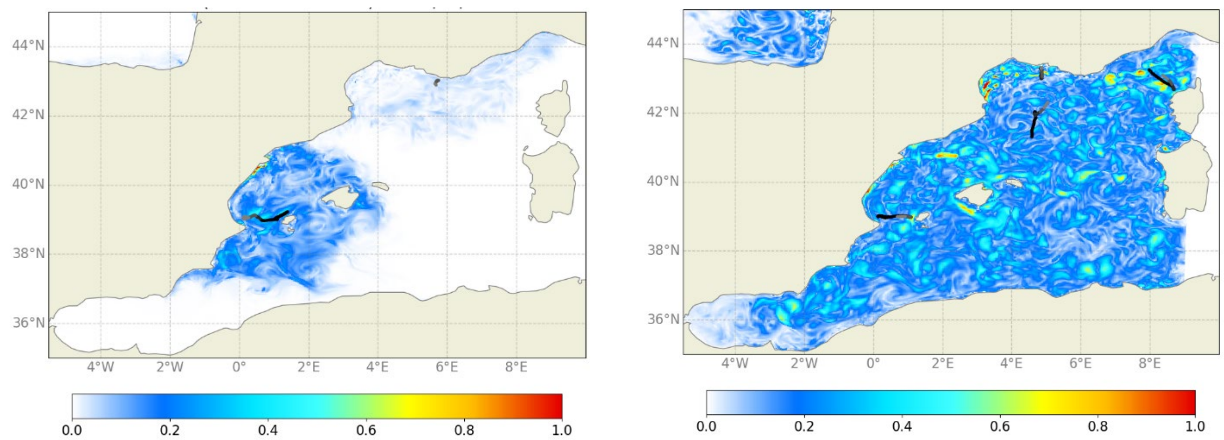

Figure 10. RMSD of temperature above 200m between NOGLID and GLIDER for two different dates: in the first cycle (2017/01/07, left) and after a month (2017/02/04, right). The gliders available for the cycle are indicated in shades of grey (shade correspond to distance in time to the date of the plot, darker means closer)

A first look into the differences between the NOGLID and GLIDER runs is presented in Figure 10. The RMSD between the two runs on the first date show the impact of the added gliders observations (in black). This is especially seen in the Balearic region where there was the first transect of the Ibiza channel recurring line. Another glider sailing from the French coast near Marseille also results in some differences between the runs along the coastline but smaller as the gliders leaves later and thus farther away in time to this particular date. After a month of simulation, the two runs diverged too much from each other from the different observations and the intrinsic chaotic variability of the system, to be able to evaluate the impact of the added observation from this diagnostic. The whole domain shows differences of mesoscale nature. This is expected as the runs should not diverge too much away from the observations apart from some phase shift of the mesoscale.

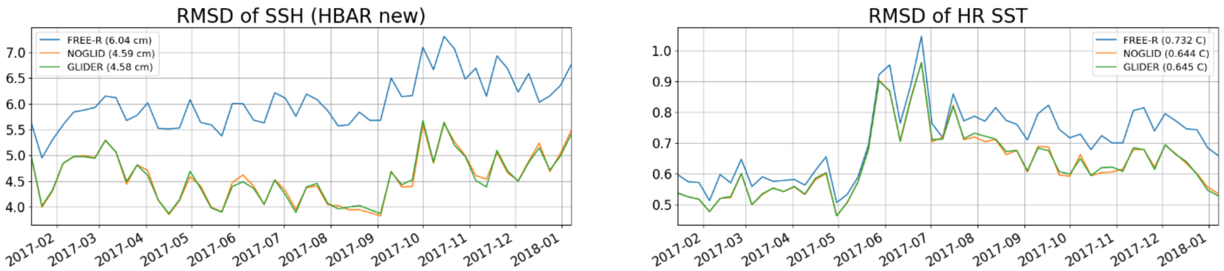

Figure 11. Evolution of RMSD compared to observations in the forecast cycle for SLA (left) and SST (right) for the different runs

Comparisons with observations are shown in Figure 11 and Figure 12 For both figures, the observation is compared to the model value at the observation time in the forecast cycle meaning before that observations were assimilated for this cycle. The impact of glider assimilation is not significant on SSH and SST over the whole domain. Concerning profiles of temperature and salinity instead, we can see some impact on the vertical view of the RMSD and bias. For the temperature, the impact is quite small but the green line is always to the left of the orange line for the RMSD. In salinity, a positive effect of the gliders is found mostly on the first 50 meters of the RMSD and reduced bias on the first 150 meters.

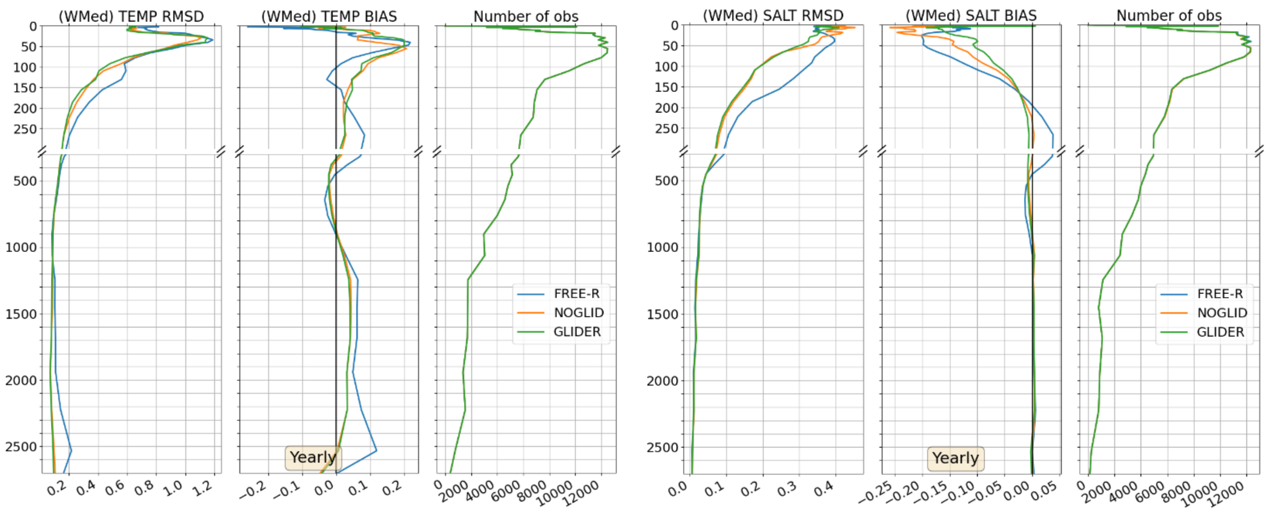

Figure 12. Vertical structure of RMSD for temperature (left) and salinity (right) compared to observed profiles in the forecast cycle for the different runs. Only profiles in the Western Mediterranean Sea are used.

Study in the Ibiza channel

Since general statistics were not completely conclusive and to better assess the impact of the assimilation of gliders in the dynamics of our simulation, we decided to focus on the Ibiza channel. Indeed, in this channel there is a recurring campaign done by SOCIB with gliders, the endurance line, aiming at monitoring the transport in this key region. On Figure 11, the section measured by these gliders is highlighted and we can clearly see how well sampled this section is during the year: at least 5 sections every two months. This will allow us to look at the ability of the model to reproduce the seasonal dynamic of a complex region.

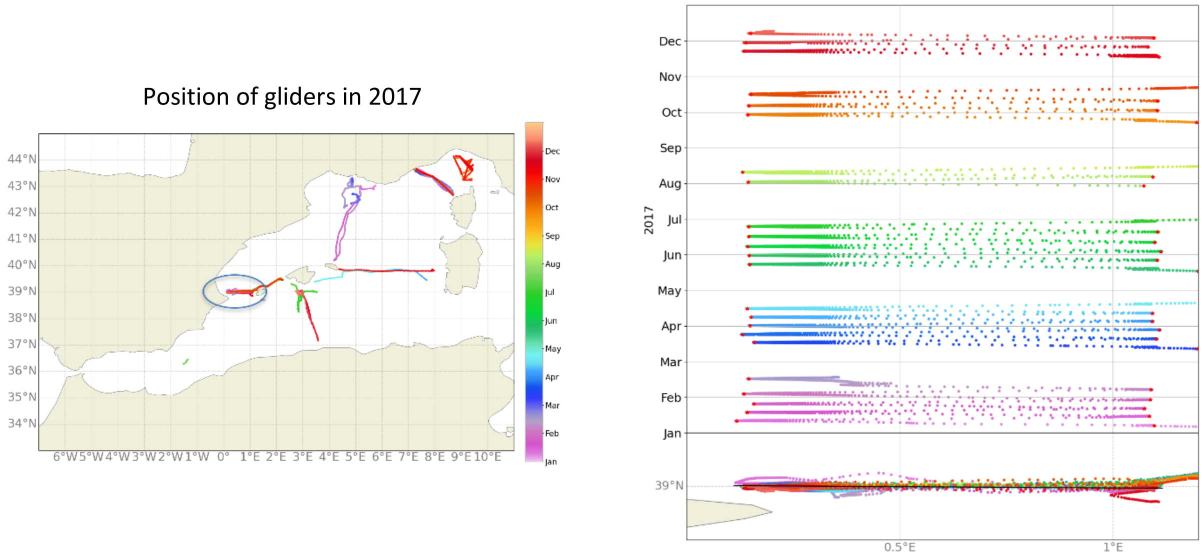

Figure 13. Position in time and space of the gliders on the Ibiza channel for the year 2017.

To assess the transports through the channel from the observations and compare it to the model estimations we will follow a similar method as described in Heslop et al. 2012. First, the observed section values of temperature and salinity are interpolated vertically and horizontally on a standard grid of 5 meters by 2 km respectively. Then we compute dynamic height from these interpolated sections by integrating the density anomalies from 800 m to 20 m. When the depth is shallower than 800m, the bottom value of the dynamic height is taken as the value of the neighbouring point and then integrated from there. The dynamic height is then smoothed horizontally with a low pass filter of 10 km. From the smoothed dynamic height, geostrophic velocities can be inferred. Figure 14 shows the geostrophic velocities obtained from the observed glider profiles on all the 44 sections that were sampled during the year 2017.

This method is then applied in the exact same manner to the model values of temperature and salinity at the glider observation points. This allows a fair comparison between the geostrophic velocities of the models and the observations. To perform this comparison on a seasonal view, bimonthly means of the geostrophic velocities are computed and presented in Figure 15. The free run performs somewhat poorly at reproducing the patterns of transport through the channel with inverted transport in January-February, and September-October or the too small subsurface currents un summer. Already what is interesting is that the assimilation of SST, SLA and T/S profiles without the gliders starts to correct the behaviour. In September-October, the current is now northward as in the observations. For the other months, some improvement is also seen such as the southward subsurface current in summer. Now, when glider assimilation is added (experiment GLIDER), the transports are getting even closer to what is obtained from the observations. The inverted circulation in January-February is significantly decreased and all other patterns visually represent a better estimation of the observed currents. This is confirmed with all RMSD of the mean sections that decrease when compared to the observations except for May-June where the surface currents RMSD is slightly better without the glider assimilation.

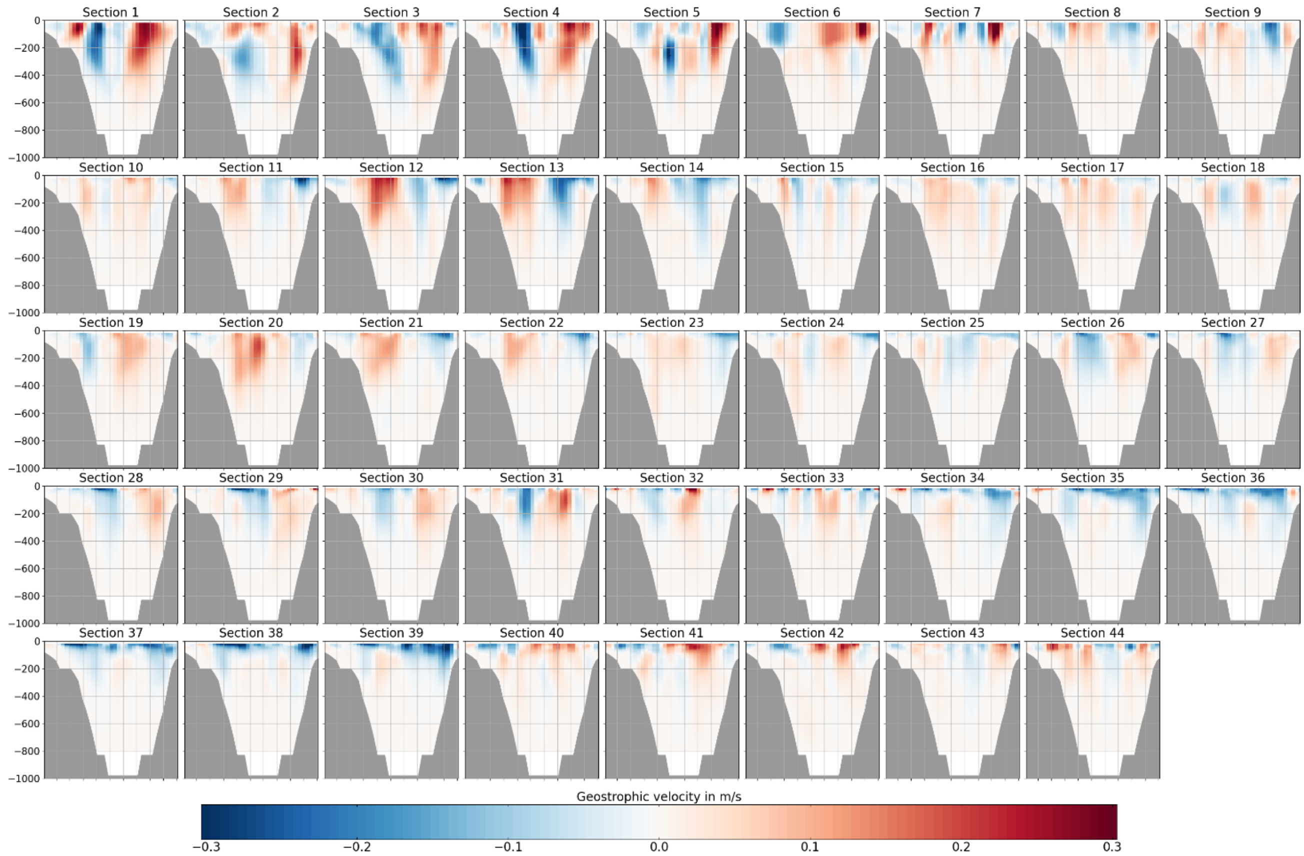

Figure 14. Geostrophic velocities computed for each glider section from the observed temperature and salinity.

From here we can compute the northward and southward transport. The results are presented in Figure 16. They confirm the positive impact of the assimilation of gliders for all RMSD estimations. The red line, representing the geostrophic transport for the GLIDER experiment, is always closer to the blue line representing the observed transport. The net transport is especially improved with the assimilation of gliders with the positive transport in spring changed to a negative transport. This gives confidence that the assimilation of gliders improves the dynamic in the region and the transport of the water masses and will impact both the Balearic Sea and the rest of the Western Mediterranean Sea.

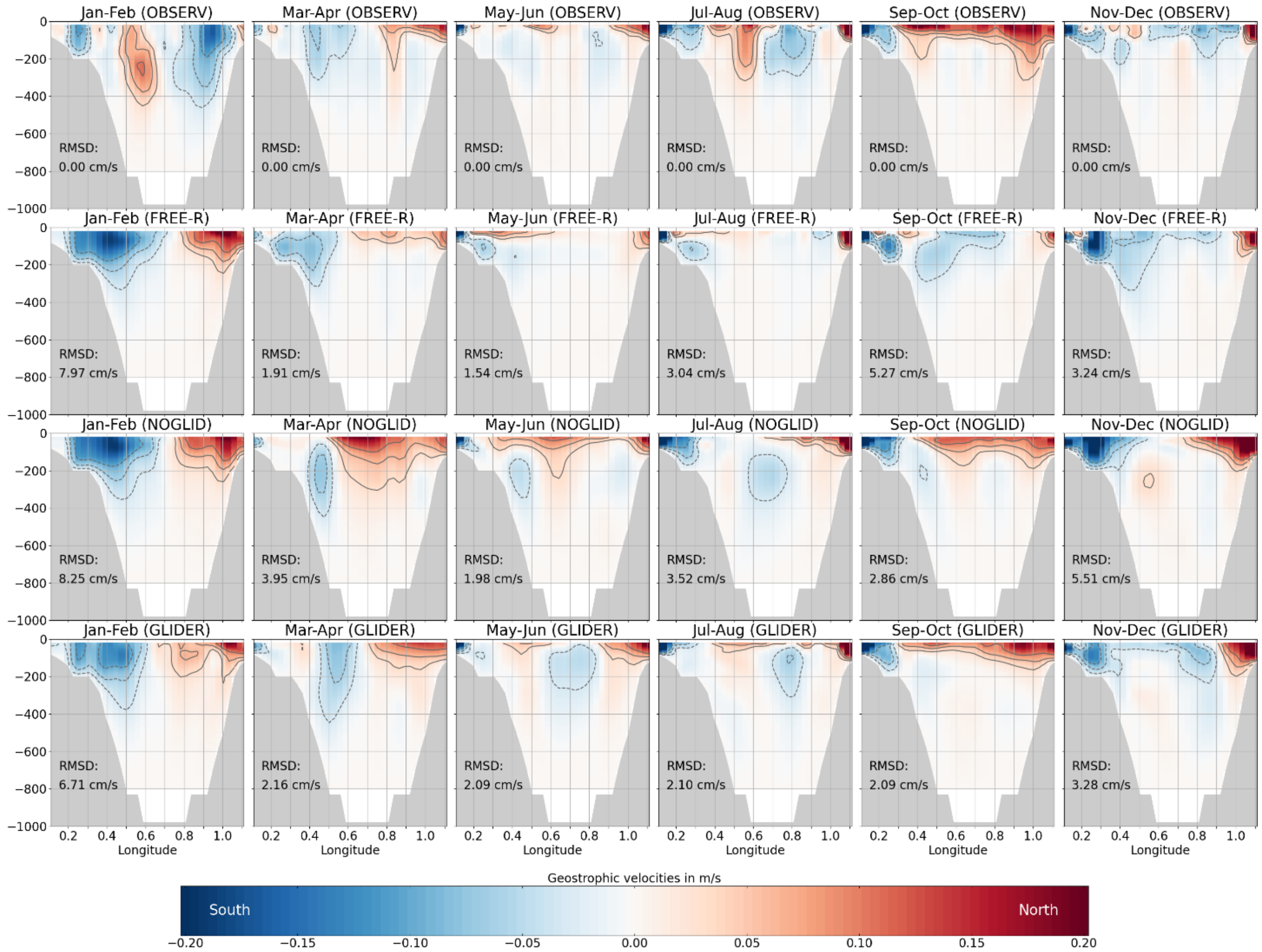

Figure 15. Sections of bimonthly means of the geostrophic velocities for the different runs and the observations.The values obtained from the gliders are in the first row while the model results are show in the 3 other rows. The RMSD between the observation and model values are indicated on each tile.

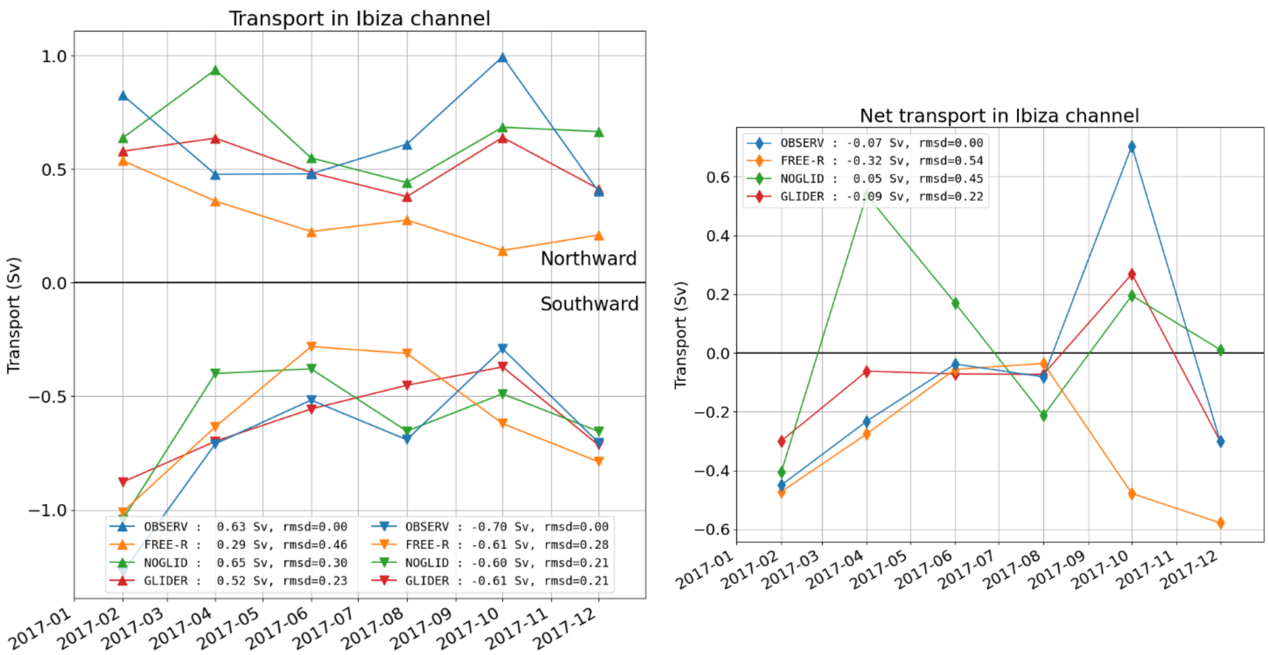

Figure 16. Bimonthly values of the geostrophic transport through the Ibiza channel for the glider observations and the three model runs.

3.3. Conclusion

In order to improve the models in the coastal area, the inclusion of new types of dataset into our assimilation systems are crucial. Multiple glider missions have successfully been launched in the last decade and provide very valuable, high-resolution observations of temperature and salinity over some sections. This study tested the use of these dataset into the operational regional system IBI over a period of one year. These observations are ingested the same way that other profiles enter the system with a modification of the observation error to account for the higher density of profiles. The results show a quite small impact over the whole domain in global statistics but when we look at local dynamic and transport in the well observed Ibiza channel, we showed that the impact of these observation is significant and these changes of transport will impact the circulation of the region.

Intercomparison (link with Task 4.2)

As stated before, an effort to make intercomparison with the experiments from task 4.2 was accomplished. We used the same glider observation dataset and on the same period. We have started to make some common diagnostics to be able to compare the impact of gliders on different systems with different strategies for the assimilation of these relatively new dataset, at least for the field of data assimilation. These results are still preliminary and will not be presented here. A common article is planned to be written.

4. CMEMS Biogeochemical model optimization using BGC-Argo float data

In this task, a data assimilation method has been developed to optimize the parameters of a 1D configuration of PISCES (Pelagic Interactions Scheme for Carbon and Ecosystem Studies), the biogeochemical model utilized by Mercator Ocean International for the Copernicus Marine Service. This method utilizes a particle filter algorithm (Mattern, Dowd, and Fennel 2013) adapted for biogeochemical systems and incorporates in situ observations collected by a BGC-Argo float in the North Atlantic, an area that experiences significant phytoplankton blooms that play a critical role in the oceanic carbon cycle.

4.1. BGC-Argo data

The data used in this study were collected from a BGC-Argo float deployed in the North Atlantic and identified by the WMO number 5904479, as shown in Figure 17. This float measures the temperature, salinity, dissolved oxygen, chlorophyll-A, BBP and nitrate concentration. To ensure the validity of our one-dimensional (1D) simulations, we carefully selected a time period during which the float was in a homogeneous water mass, indicating the absence of lateral advection. We visually analysed the time series of observed potential temperature, salinity, and potential density, as shown in Figure 18, to verify the homogeneity of the water mass. Based on our analysis, we determined that the float was in a homogeneous water mass between October 2014 (indicated by a vertical pink solid line) and January 2016. We chose this time period to constrain the 1D PISCES simulations to ensure that the 1D assumption was met.

The BGC-Argo data was complemented with pseudo-observations of nitrate, phosphate, silicate, and dissolved inorganic carbon concentrations using the CANYON-B neural network (Bittig et al. 2018). The CANYON-B algorithm estimates the vertical profiles of nutrients and the carbonate system variables by utilizing the float pressure, temperature, salinity, and oxygen measurements along with the associated geolocalization and sampling date data.

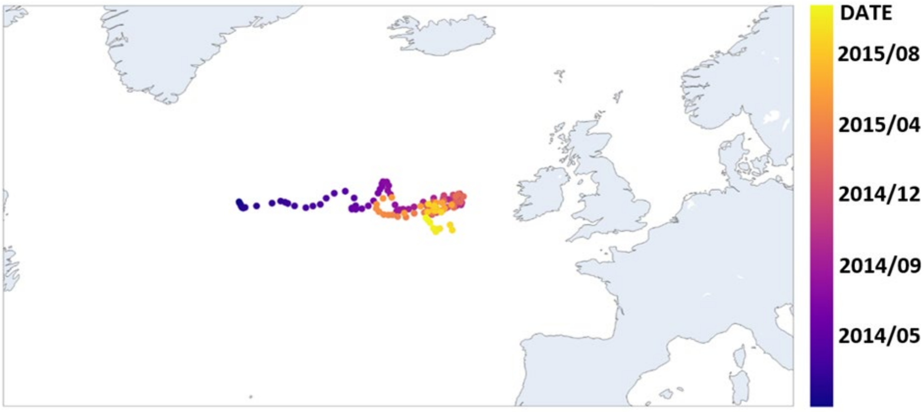

Figure 17. Trajectory of the BGC-Argo float with WMO number 5904479 in the North Atlantic. The colors represent the time of sampling of each vertical profile acquired by the float, with warmer colors indicating more recent profiles and cooler colors indicating older profiles. The float sampled the water column every 10 days from the surface down to 2000 meters, measuring physical and biogeochemical variables such as temperature, salinity, oxygen, nitrate, chlorophyll, and suspended particles. The trajectory shown here covers a period of almost two years, from February 2014 to December 2015.

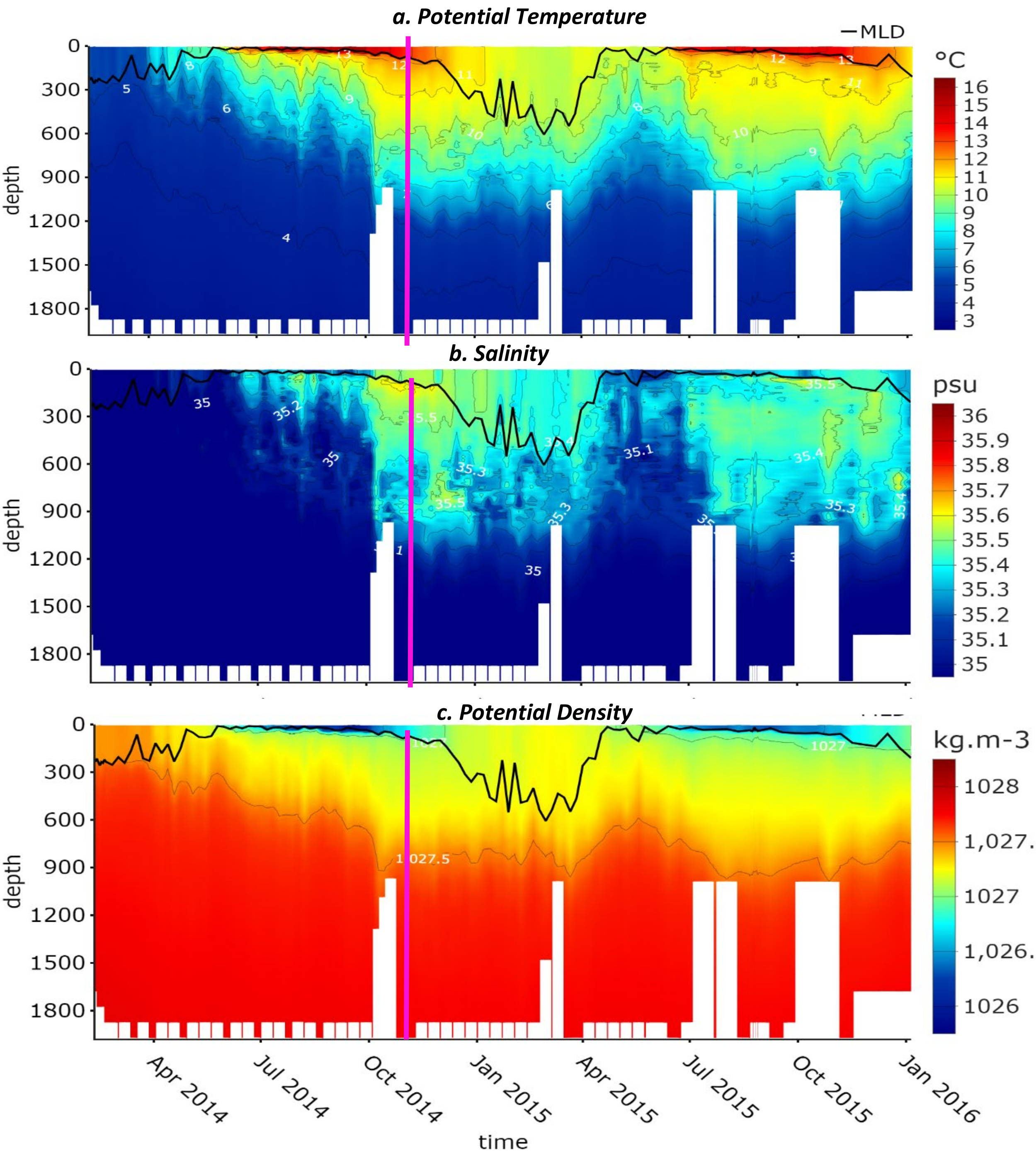

Figure 18. Temporal evolution of vertical profiles of potential temperature, salinity, and potential density (panels a, b, and c, respectively) observed by the float 5904479. The depth of the mixed layer (MLD) is indicated by the solid black line, and the water mass change is represented by the vertical solid pink line.

4.2. Methods

Biogeochemical model PISCES

PISCES (Aumont et al. 2015) simulates the lower trophic levels of the marine ecosystem, including phytoplankton and zooplankton, as well as the biogeochemical cycles of carbon and major nutrients. PISCES is designed to be used at spatial scales ranging from regional to global and temporal scales ranging from short-term (seasonal) to long-term (climate change, paleoclimate). The model has 24 prognostic variables grouped into four main compartments: nutrients, phytoplankton, zooplankton, and organic matter. There are five limiting nutrients (nitrate, ammonium, phosphate, silicate, and iron) for phytoplankton growth. Phytoplankton and zooplankton are modeled in two different size groups, and the organic matter compartment includes dissolved and particulate matter. Phytoplankton uses solar radiation and nutrients for photosynthesis, while zooplankton feeds on phytoplankton and organic matter. Dead organisms are decomposed, returning to an inorganic state and becoming available again for phytoplankton. In addition, PISCES simulates dissolved inorganic carbon, alkalinity, and dissolved oxygen, providing a comprehensive understanding of the marine ecosystem and biogeochemical processes.

1D Configuration of PISCES

To generate a large number of simulations while maintaining reasonable computation time, we coupled NEMO and PISCES in offline mode using a one-dimensional configuration (Reffray, Bourdalle-Badie, and Calone 2015). In this approach, PISCES 1D uses previously computed physical outputs from the global ocean physical analysis and forecasting system of the Copernicus Marine Service[2]. The one-dimensional modelling approach assumes a vertical profile of the water column, where horizontal exchanges are negligible due to the absence of circulation. This means that there is no horizontal advection and horizontal mixing. The conservation of momentum also results in the absence of vertical advection. The vertical processes are solely dependent on turbulence, atmospheric conditions at the surface, and solar penetration. Consequently, we can express the temporal evolution of a biogeochemical tracer C using the following equation:

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -\frac{\mathrm{d}\left(k_z \frac{dC}{dz}\right)}{\mathrm{d}z} + (Gains - Losses)(C) \qquad (1)$$

In this equation, the first term on the right-hand side represents the vertical mixing, where k_z is the vertical diffusion coefficient. The second term, (Gains-Losses) (C), represents the balance of gains and losses of the tracer C due to the biogeochemical processes affecting it.

To initialize the PISCES simulation, the 24 prognostic variables are required as inputs. In this study, we utilized in-situ observations from the BGC-Argo float, including chlorophyll, nitrate, phosphate, silicate, oxygen, particulate organic carbon, and dissolved inorganic carbon. For the non-observed variables, we incorporated the output from the global ocean BGC analysis and forecasting system of the Copernicus Marine Service[3] (hereinafter denoted PISCES 3D). Additionally, the 1D NEMO-PISCES model requires knowledge of the physical environmental conditions along the float's trajectory. To obtain these conditions, we extracted daily dynamical forcings such as temperature, salinity, and vertical diffusion coefficient from the global ocean

---

[2] https://data.marine.copernicus.eu/product/GLOBAL_ANALYSISFORECAST_PHY_001_024

[3] https://data.marine.copernicus.eu/product/GLOBAL_ANALYSIS_FORECAST_BIO_001_028

physical analysis and forecasting system of the Copernicus Marine Service. Finally, the 1D configuration incorporates the same atmospheric forcings at the ocean surface as the operational circulation and biogeochemical models mentioned above, including wind, solar flux, Photosynthetically Available Radiation (PAR), precipitation-evaporation balance (P-E), and atmospheric nutrient inputs.

4.3. Results and discussion

The time series of each state variable averaged in the mixed layer along the trajectories of the float are shown in Figure 19. The red curves represent the outputs from PISCES 1D using the reference set of parameters, while the orange curves represent the outputs from PISCES 3D. The blue lines represent the BGC-Argo observations, with error bars indicating both observations error and model representativeness error. The grey lines represent the ensemble members, and the black line represents the ensemble member with the highest weight, i.e., the ensemble member that is most likely to explain the observations (hereinafter denoted the best member).

The float sampled a region of the North Atlantic where a phytoplankton bloom occurred during the spring season, as evidenced by the time series of chlorophyll a, particulate organic carbon, nitrate, phosphate, silicate, dissolved inorganic carbon, and oxygen. A phytoplankton bloom is a rapid and substantial increase in the population of phytoplankton in a water body, resulting in a sudden increase in chlorophyll a concentration due to the rapid growth of the phytoplankton. The bloom is also characterized by an increase in the concentration of particulate organic carbon, while the concentration of dissolved inorganic carbon and nutrients decrease due to phytoplankton consumption. Oxygen levels tend to increase during the bloom due to photosynthetic activity. The spring bloom is typically followed by a decline in chlorophyll a, particulate organic carbon, and oxygen during the summer and fall. This decline is due to a decrease in light availability and associated photosynthesis. Concurrently, there is an increase in dissolved inorganic carbon and nutrients, reflecting the resupply of nutrients from deeper waters to the mixed layer.

The optimization of PISCES 1D model using the particle filter algorithm greatly improves the model's ability to reproduce the North Atlantic bloom. The optimized model is able to reproduce the seven variables more accurately than PISCES 3D and PISCES 1D with the reference set of parameters, as confirmed by observations. This improvement is confirmed by a statistical evaluation using the Root Mean Square Error (RMSE). Table 2 shows that the RMSE values between the observations and the best member are the lowest for all variables. Furthermore, Table 3 shows that the optimized version of PISCES 1D showed a significant decrease in the RMSEs compared to the 1D and 3D references. These statistical results validate the visual inspection of the time series and confirm the conclusion that the optimization of the PISCES 1D model using the particle filter algorithm has significantly improved the model's ability to reproduce the North Atlantic bloom.

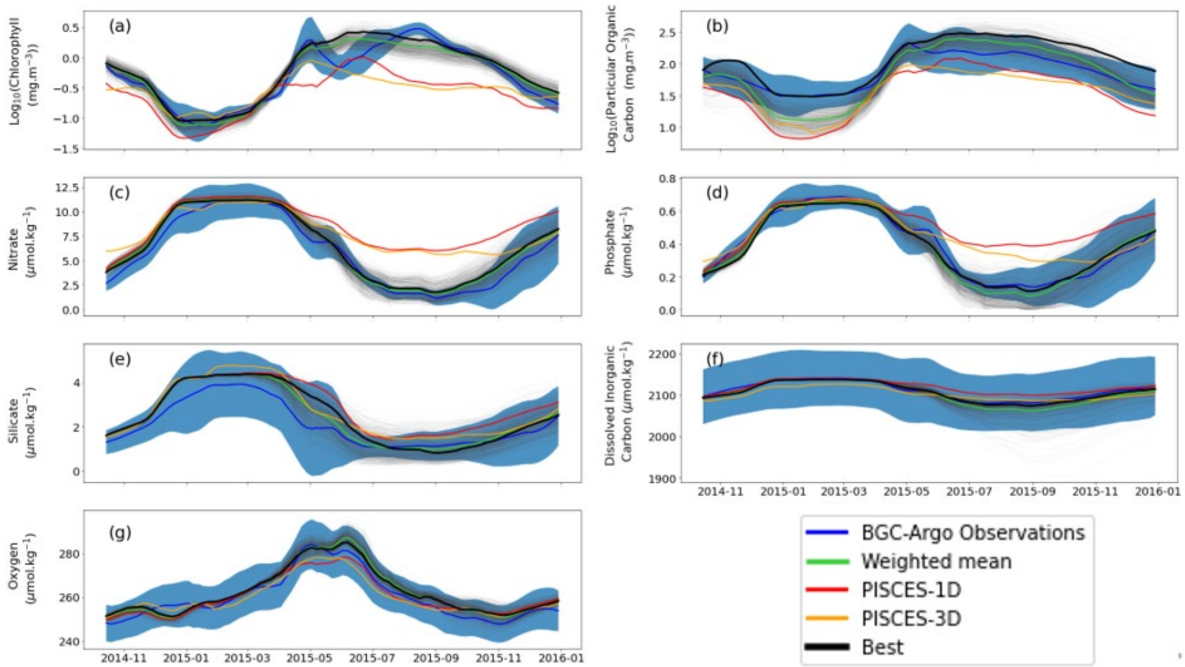

*Figure 19. Time series of chlorophyll a **(a)**, particulate organic carbon **(b)**, nitrate **(c)**, phosphate **(d)**, silicate **(e)**, dissolved inorganic carbon **(f)**, and oxygen **(g)** averaged in the mixed layer along the trajectory of the BGC-Argo. The blue line represents the BGC-Argo observations, and the blue shading the observational errors. The red line depicts the time series of the reference 1D PISCES model simulation. The orange line represents the PISCES 3D, and the grey lines depict all the ensemble members of the particle filter algorithm. Finally, the black line shows the best member of the particle filter algorithm and the green line represent the weighted mean.*

Table 2. Root Mean Square Error (RMSE) Comparison of Chlorophyll-a, Nitrate, Phosphate, Silicate, Dissolved Inorganic Carbon, Particulate Organic Carbon, and Oxygen Between the Best Member and PISCES 1D/3D.

RMSE / Variables	Best	PISCES 1D	PISCES 3D
log10(Chlorophyll-a ($mg.m^{-3}$))	0.17	0.41	0.41
Nitrate ($\mu mol.kg^{-1}$)	0.85	3.07	2.53
Phosphate ($\mu mol.kg^{-1}$)	0.04	0.15	0.09
Silicate ($\mu mol.kg^{-1}$)	0.61	0.87	0.65
Dissolved inorganic carbon ($\mu mol.kg^{-1}$)	5.38	11.9	12.85
log10(Particulate organic carbon ($mg.m^{-3}$))	0.23	0.37	0.33
Oxygen ($\mu mol.kg^{-1}$)	3.18	3.44	3.17

Table 3. Comparison of percent RMSE differences between the best member and PISCES 1D/3D for chlorophyll-a, nitrate, phosphate, silicate, dissolved inorganic carbon, particulate organic carbon, and oxygen.

% RMSE difference / Variables	PISCES 1D	PISCES 3D
log10(Chlorophyll-a)	-59 %	-59 %
Nitrate	-72 %	- 66 %
Phosphate	-74 %	-55 %
Silicate	-30 %	-6 %
Dissolved inorganic carbon	-55 %	-58 %
log10(Particulate organic carbon)	-39 %	-30 %
Oxygen	-8 %	0

4.4. Conclusions and perspective

In this task, we have demonstrated the effectiveness of a particle filter algorithm in assimilating BGC-Argo observations and optimizing a 1D configuration of PISCES in the North Atlantic. The optimization of the PISCES 1D model using the particle filter algorithm significantly improves the model's ability to reproduce the North Atlantic bloom. In particular, the optimized model can reproduce the time series of chlorophyll, particulate organic carbon, nitrate, phosphate, silicate, dissolved inorganic carbon, and oxygen more accurately than both the 1D version of PISCES with the reference set of parameters and the global ocean BGC analysis and forecasting system of the Copernicus Marine Service (European Union-Copernicus Marine Service 2019). Furthermore, the results of an OSSE confirm the effectiveness of the particle filter algorithm in assimilating observations and providing accurate estimation of model parameters. Our results suggest that this approach can be applied to other oceanic regions and used to improve the accuracy and reliability of BGC models for monitoring and forecasting purposes.

A PhD student is currently continuing this research, adapting the methodology developed in the framework of EUROSEA for global application. The ultimate goal is to derive a spatial map of optimized parameters and implement it in the global ocean BGC analysis and forecasting system of the Copernicus Marine Service. This would provide more accurate and reliable predictions of marine ecosystems on a global scale, thereby helping to better understand and manage the health of our oceans.

5. Conclusion

In this report, we analyse the use and the efficiency of the in-situ observations to constrain regional and global Mercator Ocean systems. Physical and biogeochemical variables are considered. The in-situ observations are used either to estimate physical ocean state at global and regional scale via data assimilation or to estimate BGC model parameters.

The impact of the physical in situ observations assimilated in open ocean and coastal areas is assessed with numerical data assimilation experiments. The experiments are conducted with the regional 1/36° resolution and global 1/12° resolution systems operated by Mercator Ocean for the Copernicus Marine Service. For the global physical ocean, the focus is on the tropical ocean to better understand how the tropical mooring observations constrain the intraseasonal to daily variability and the complementarity with satellite observations.

The tropical moorings provide unique high frequency observations at different depth, but they are far away from each other, so part of the signal in the observation are decorrelated from one mooring to the others. It is only via an integrated approach, as data assimilation into a dynamical model and complementarity with other observing networks that those observations can efficiently constrain the tropical ocean circulation. The spectral analysis of the Glorys 1/12° global reanalysis clearly shows a great improvement of the representativity of the tropical dynamic, dominated by propagating waves, when observations are assimilated compared to a simulation without assimilation. Not only the variability but also the phase of the tropical waves is improved, this is the case from intra seasonal to daily time scales. As the satellite observations brings higher spatial resolution between the tropical moorings but for the ocean surface, the tropical mooring and Argo profiles constrain the larger scale ocean thermohaline vertical structure (EuroSea D2.2; Gasparin et al., 2023).

The ocean below 2000 m depth is still largely under constrained as very few observations exist. Some deep ocean basins, as the Antarctic deep ocean, shows significant trend over the past decade but they are still not accurately monitored. Four CMEMS reanalysis based on different data assimilation schemes and model configuration set up were considered to assess the uncertainty of the deep ocean temperature and salinity fields. An intercomparison was conducted against each of the four reanalysis and deep Argo observations. Based on the spread of those four deep ocean reanalysis estimates, large uncertainties were estimated in representing local heat and freshwater content in the deep ocean. Additionally, temperature and salinity field comparison with deep Argo observations demonstrates that reanalysis errors in the deep ocean are of the same size as or even stronger than the observed deep ocean signal. The sparsity of deep ocean observations prevents to have reliable estimate of the deep ocean mean and variability. OSSE suggested that the deployment of a global deep Argo array will significantly constrain the deep ocean in reanalysis to be closer to the observations (Gasparin et al., 2020).

At regional and coastal scales, the ocean circulation is dominated by higher frequency, smaller scale processes than the open ocean which requires different observation strategy to be well monitor. The shelf regions are more responsive to high frequency atmospheric forcing by the wind and pressure, in addition to be forced by tides. The impact of assimilating high frequency and high-resolution observations provided by gliders on European shelves is analysed with the regional Iberic Biscay and Irish (IBI) system. Two types of glider deployment were considered: one "one time" line in the Atlantic Ocean and repeated lines in the western Mediterranean Sea, as between the Balearic Islands and Spain. From the comparison of data assimilation experiments with and without glider assimilation, in the latest version of the operational regional IBI system, it was shown that the repetitive glider sections help to constrain the transport of water masses flowing across those sections. The influence of the glider DA can be seen away from the observed section. The small-scale changes in the analysis and forecasts due to the glider assimilation are still difficult to evaluate due to the sparsity of in situ observations in those coastal regions compared to their high variability. The data assimilation experiments done in this task were defined jointly with the Task 4.2 were similar experiments but with a different assimilation system were conducted. This will help to draw more robust conclusions since OSE/OSSE are highly dependent on the DA system. A common article is planned to be published.

BGC ocean models are less mature than physical ocean models and some variable dependencies are still based on empirical functions. In this task, Argo BGC profile observations were used to optimize the parameters of the global CMEMS biogeochemical model, PISCES. A particle filter algorithm was chosen to optimize a 1D configuration of PISCES in the North Atlantic. The optimization of the PISCES 1D model significantly improves the model's ability to reproduce the North Atlantic bloom. In particular, the optimized model can reproduce the time series of chlorophyll, particulate organic carbon, nitrate, phosphate, silicate, dissolved inorganic carbon, and oxygen more accurately than both the 1D version of PISCES with the reference set of parameters and the global ocean BGC analysis and forecasting system of the Copernicus Marine Service (European Union-Copernicus Marine Service 2019). Since this approach is validated, the ultimate goal is to apply the same method to other BGC Argo floats at global scale to derive a spatial map of optimized parameters and implement it in the global ocean BGC analysis and forecasting system of the Copernicus Marine Service. This would provide more accurate and reliable predictions of marine ecosystems on a global scale, thereby helping to better understand and manage the health of our oceans.

Recommendations on the in-situ network extensions for real time ocean monitoring are given based on those results, and the one also obtained in the WP2, Task 2.2 where data assimilation experiments but with simulated observations where conducted. Argo extension and the complementarity with satellite altimetry was extensively studied.

The deep ocean (below 2000 meter) is impacted by climate change and should be considered for earth energy imbalance monitoring, heat storage and sea level rise computation. At present, Copernicus global ocean reanalysis show uncertainty in the order of the climate signal. Extending the Argo array to the deep ocean would help to better constrain the deep ocean variability and trend in ocean analysis and evaluate the associated uncertainties.

From the Glorys12 and twin experiment without assimilation, we showed that the tropical ocean variability down to a day, is well represented. The assimilation of satellite observations, SST and SLA, sets the surface conditions as the tropical mooring profile assimilation sets the vertical structure that drives the propagation properties of the waves. Long time series of mooring observations are very well suited to evaluate the realism of simulation from high frequency hourly variability, as the diurnal cycle, up to the interannual variability. Enhanced resolution of the vertical profiles especially at the thermocline depth, and addition of instruments such as current meter at different depth, will be important for validation and to better constrain the variability and strength of the thermocline and halocline at all temporal scales in real time global monitoring system and ocean reanalysis.

On the shelf, the in-situ observation coverage is quite sparse compared to spatial and temporal scales of variability. Increased in situ observation coverage would help to evaluate developments made in regional and coastal model and constrain them via data assimilation. From the glider data assimilation experiments, we shown that a repetitive sampling is a good strategy to maintain a coherent constrain in time. Close to the coast and on the shelf, ocean currents are very important to be well estimated since they strongly influence coastal zone by advecting freshwater from rivers, heat and BGC properties. Expanding the capacity of direct measurement of ocean current close to the coast (with HF radar for example) would be important.

The BGC models and observations still need to progress. A better sampling of the ocean with in-situ BGC observations will help to evaluate and tune the BGC model, as well as complement the satellite BGC observations (ocean colour) that are today assimilated in BGC operational systems. This should improve the estimation of quantities of interest for CMEMS as the air/sea fluxes of CO_2, acidification, deoxygenation and primary production.

For all variables and regions, the complementarity between in situ physical and BGC observations and satellite observations, that mostly provide surface higher resolution observation, should be considered. As shown in WP2-T2.2, the assimilated satellite observations are efficiently constraining the global ocean physical circulation at intermediate scale as the in-situ observations are constraining the large-scale variability (Gasparin et al., 2023) and vertical structure of the ocean needed to well project the satellite surface observations onto the water column.

More recommendations from different Copernicus marine Monitoring and Forecasting Centres as well as Thematic Data assembly centres can be found in the document “CMEMS requirements for the evolution of the Copernicus In Situ Component”.

References

Aumont, O., C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen. 2015. 'PISCES-v2: An Ocean Biogeochemical Model for Carbon and Ecosystem Studies'. *Geoscientific Model Development* 8 (8): 2465–2513. https://doi.org/10.5194/gmd-8-2465-2015.

Bittig, Henry C., Tobias Steinhoff, Hervé Claustre, Björn Fiedler, Nancy L. Williams, Raphaëlle Sauzède, Arne Körtzinger, and Jean-Pierre Gattuso. 2018. 'An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO2 Variables and Nutrient Concentrations from T, S, and O2 Data Using Bayesian Neural Networks'. *Frontiers in Marine Science* 5: 328.

European Union-Copernicus Marine Service. 2019. 'Global Ocean Biogeochemistry Analysis and Forecast'. Mercator Ocean International. https://doi.org/10.48670/MOI-00015.

Gasparin, F., Hamon M., Remy E., Le Traon, P.-Y., 2020. How Deep Argo Will Improve the Deep Ocean in an Ocean Reanalysis. Journal of Climate. 33. 10.1175/JCLI-D-19-0208.1.

Gasparin F, Lellouche J-M, Cravatte SE, Ruggiero G, Rohith B, Le Traon PY and Rémy E. 2023. On the control of spatial and temporal oceanic scales by existing and future observing systems: An observing system simulation experiment approach. Front. Mar. Sci. 10:1021650. doi: 10.3389/fmars.2023.1021650.

Lellouche J.-M., E. Greiner, R. Bourdallé-Badie, G. Garric, A. Melet, M. Drévillon, C. Bricaud, M. Hamon, O. Le Galloudec, C. Regnier, T. Candela, C.-E. Testut, F. Gasparin, G. Ruggiero, M. Benkiran, Y. Drillet, P.-Y. Le Traon. 2021. The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis. Frontiers in Earth Sci., Volume 9 - 2021, https://doi.org/10.3389/feart.2021.698876

Mattern, Jann Paul, Michael Dowd, and Katja Fennel. 2013. Particle Filter-Based Data Assimilation for a Three-Dimensional Biological Ocean Model and Satellite Observations. Journal of Geophysical Research: Oceans 118 (5): 2746–60. https://doi.org/10.1002/jgrc.20213.

Mercator Ocean, EUROGOOS, and CMEMS partners. 2021. Copernicus Marine Service requirements for the evolution of the Copernicus In Situ Component

Reffray, G., R. Bourdalle-Badie, and C. Calone. 2015. 'Modelling Turbulent Vertical Mixing Sensitivity Using a 1-D Version of NEMO'. Geoscientific Model Development 8 (1): 69–86. https://doi.org/10.5194/gmd-8-69-2015.

Turpin, V., Remy, E., and Le Traon, P. Y.: How essential are Argo observations to constrain a global ocean data assimilation system?, Ocean Sci., 12, 257–274, https://doi.org/10.5194/os-12-257-2016, 2016.