Dense CTD survey versus glider fleet sampling: comparison of the performance for regional ocean prediction West of Sardinia

Jaime Hernandez-Lasheras[1] and Baptiste Mourre[1]

[1]Balearic Islands Coastal Observing and Forecasting System - SOCIB

Correspondence to: Jaime Hernandez (jhernandez@socib.es)

Abstract.

The REP14-MED sea trial carried out off the West coast of Sardinia in June 2014 provided a rich set of observations from both ship-based CTDs and a fleet of underwater gliders. We present the results of several simulations assimilating data either from CTDs or from different subsets of glider data, including up to 8 vehicles, in addition to satellite sea level anomalies, surface temperature and Argo profiles. The WMOP regional ocean model is used with a Local Mutimodel Ensemble Optimal Interpolation scheme to recursively ingest both lower-resolution large scale and dense local observations over the whole sea trial duration. Results show the capacity of the system to ingest both type of data, leading to improvements in the representation of all assimilated variables. These improvements persist during the 3-day periods separating two analysis. At the same time, the system presents some limitations in properly representing the smaller scale structures, which are smoothed out by the model error covariances provided by the ensemble. An evaluation of the forecasts using independent measurements from shipborne CTDs and a towed Scanfish deployed at the end of the sea trial shows that the simulations assimilating initial CTD data reduce the error by 39% on average with respect to the simulation without data assimilation. In the glider-data-assimilative experiments, the forecast error is reduced as the number of vehicles increases. The simulation assimilating CTDs outperforms the simulations assimilating data from one to four gliders. A fleet of eight gliders provides a similar performance as the 10-km spaced CTD initilization survey in these experiments, with an overall 40% model error reduction capacity with respect to the simulation without data assimilation when comparing against independent campaign observations.

1 Introduction

Short-term regional ocean prediction is important to respond to maritime emergencies related to search-and-rescue or accidental contamination, for maritime security or as a support to naval operations. High resolution regional ocean circulation models are used to downscale the conditions provided by operational large scale models, so as to represent mesoscale and coastal processes which are not properly resolved in the large scale simulations but play a major role in ocean transports of relevance for practical applications. Data assimilation (DA), which aims at optimally combining dynamical ocean models with in-situ and remotely sensed observations, constitutes an essential component of the prediction systems since it helps to recursively improve the initial conditions used for the prediction phases.

In order to constrain errors and remain as close as possible to reality, models must be fed with different kinds of observations. Satellites play a key role, providing regular near real time data of surface variables such as temperature and sea surface height. Water column measurements are more scarce. The Argo program provides routine temperature and salinity profiles at regular intervals, but the distance between floats is insufficient to monitor the mesoscale and finer scale variability.

5 Dedicated campaigns providing underwater measurements from ship data or glider measurements provide complementary data over specific areas. Efficient DA systems should be able to advantageously combine large scale observations over a large domain with more dense, high resolution observations in specific areas. Traditionally, campaigns onboard research-vessels have been carried out to collect dense CTD data to initialize regional ocean prediction systems. However, campaigns are not always possible. They depend on ship availability, weather, access to the area of interest and they remain very expensive. Recent evo-

10 lutions in technology allow to deploy autonomous underwater vehicles (AUV) such as gliders to collect dense hydrographic data over specific areas of interest (Testor et al., 2010; Ruiz et al., 2012; Rudnick, 2016; Liblik et al., 2016). Gliders are able to operate under hard maritime situations and to reach difficult access areas, with an overall cost reduced compared to traditional ship campaigns. Glider missions are typically planned to reach a series of locations commonly called waypoints, in order to track areas of interest and adapt to safety conditions (Garau et al., 2009). Their controllability also permits adaptive sampling

15 procedures, changing their route along the mission with the objective to collect data at optimal locations to maximize their information content (e.g. Lermusiaux (2007); Mourre and Alvarez (2012)).

The potential of gliders to sample fine scale processes and to identify different water masses has been demonstrated (Pascual et al., 2017; Heslop et al., 2012), as well as their capability to improve ocean model predictions via DA (e.g. Melet et al. (2012); Shulman et al. (2008); Pan et al. (2014); Mourre and Chiggiato (2014)). The question arises whether the sampling offered by

20 a fleet of several gliders is as useful as a traditional ship-based CTD survey for regional ocean forecasting applications. This is the question we are addressing in this paper.

The rich dataset collected during the REP14-MED campaign is used for this purpose. REP14-MED took place in June 2014 offshore the Western coast of Sardinia (Onken et al., 2018; Knoll et al., 2017). Two research vessels tracked in parallel a 100 x 100 km area during a 20-day period, providing dense CTD sampling with a 10-km separation and continuous towed CTD

25 measurements for limited periods of time. In addition, a fleet of 11 AUVs were deployed performing back-and forth sections perpendicular to the coast with a 10km vehicle intertrack distance.

The REP14-MED experiment area took place in the so-called Algero-Provençal basin of the Western Mediterranean Sea, one of the most dynamic areas of the entire Mediterranean sea (Olita et al., 2011; Millot, 1999). The region is strongly influenced by instabilities of the Algerian current, which generate intense anticyclonic eddies which can propagate northward towards the

30 western Sardinian coast (Robinson et al., 2001; Testor et al., 2005; Escudier et al., 2016a). Such eddies can last from weeks to months. They are responsible for an intense mesoscale activity in the study region (Santinelli et al., 2008). At depth, eddies are also generated from the interaction between the Algerian Gyre and inflows of Levantine Intermediate Water (LIW) and Tyrrhenian Deep Water (TDW) coming from the Channel of Sardinia (Testor et al., 2005).

The assimilation system used in this work follows an EnOI (Ensemble Optimal Interpolation) scheme (Evensen, 2003). This

35 method is widely used in oceanographic regional models as it provides a cost-effective approach when compared with more

advanced methods such as EnKF or 4Dvar (Oke et al., 2008), suitable for operational implementations. EnOI is a 3D sequential DA method. A stationary ensemble of model simulations is used to calculate background covariances. Contrary to the EnKF which requires to evolve an ensemble of simulations, a single model integration is required between two analysis steps in the EnOI, making the method numerically efficient. The EnOI provides a suboptimal solution compared to the EnKF (Sakov and Sandery, 2015). However, it represents a good alternative allowing to use a large ensemble size together with localization when necessary (Oke et al., 2007). In this work, a Local Multimodel EnOI scheme is implemented. "Mulitmodel" represents the fact that the library of ocean states is built using different long-term model simulations. "Local" means that the EnOI analysis comprises some domain localization to reduce the impact of remote observations.

The paper is organized as follows: Section 2 presents the observing and modelling frameworks, as well as the specific forecast experiment. Section 3 details the results, which are further discussed in Section 4. Finally, Section 5 concludes the paper.

2 Data and Methods

For this study, several simulations were produced assimilating different datasets from the REP14-MED campaign. This Section describes the model and data used and the methodology followed in this work.

2.1 REP14-MED experiment

The REP14-MED sea trial (Onken et al., 2018) was conducted in the framework of the EKOE (*Environmental Knowledge and Operational Effectiveness*) research program of the Centre for Maritime Research and Experimentation (CMRE, Science and Technology Organization - NATO). It is part of a series of sea trials dedicated to *Rapid Environmental Assessment* (REA), denoted by the acronym REP (*Recognized Environmental Picture*). Leaded by CMRE and supported by 20 partners, the trial took place during 20 days in June 2014, with NATO Alliance and the German Planet research vessels (RV) conducting a joint survey over an aproximately 100×100 km area off the west coast of Sardinia (figure 2). A massive amount of data was collected during the campaign with various oceanographic instruments, including CTD stations, towed Scanfish and CTD chain, ship mounted ADCP, shallow and deep underwater gliders, moorings, surface drifters and profiling floats. The sampling was divided into three legs. The time distribution of the collection of observations used in the present work is illustrated in figure 1.

During Leg 1, both RV conducted a parallel sampling of the target area, collecting CTD data with a 10km distance between stations over a 5 day period. During Leg 3, CTD data were collected with the same density, yet over a reduced spatial extension, providing very valuable data to validate the forecast experiments. Casts reach down to 1000m deep when possible. In a few profiles, they even get to greater depths in order to characterize deep water masses. Additional towed Scanfish measurements of temperature and salinity down to 200m depth allowed to complete the characterization of the area during Leg 3. At the same time, and during the whole duration of the campaign, eight gliders were considered, travelling continously along back and forth transects perpendicularly to the Sardinian coast. Five of these gliders were deep gliders submerging to depths down to 800m, the remaning three were shallow water platforms collecting data in the upper 200m only. Each of these single transects
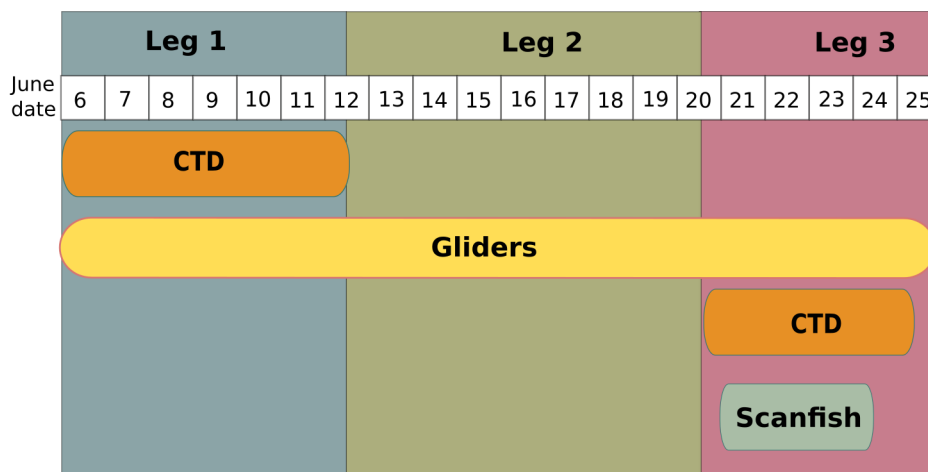
Figure 1. Sampling schedule in June 2014 of the REP14-MED sea-trial data employed in the present work.

was completed in about three days in each way. Notice that three additional gliders were deployed during the sea trial, but due to technical problems, duplication of the track and lack of processed data, they were discarded here. All glider tracks are approximately parallel to each other, with an intertrack distance around 10km, thus covering the target area. Figure 2 situates the sampling area in the Western Mediterranean Sea and shows the position of CTD, glider and Scanfish data during the

5   different legs of the sea trial.

2.2   Model and Data Assimilation System

The model used in this work is the Western Mediterranean OPerational model (WMOP, Juza et al. (2016)), covering a domain extending from Gibraltar strait to the Sardinia channel. WMOP is based on ROMS (Shchepetkin and McWilliams, 2005), a three-dimensional free-surface, sigma coordinate, primitive equations model using split-explicit time stepping with Boussi-

10   nesq and hydrostatic approximations. WMOP is set-up with 32 vertical levels and a 2km spatial resolution. It is forced at the surface using the 3-hourly and 5-km resolution HIRLAM atmospheric model provided by the Spanish meteorological agency (AEMET). In this work, the initial state of the forecasts experiments is provided by the simulation fields on 1 June 2014 of a seven-year long free run WMOP simulation spanning the period 2009-2015. This simulation uses initial state and boundary conditions from the CMEMS-MED reanalysis (Simoncelli, 2017). In addition, several WMOP free run hindcast simulations

15   were generated including modifications of the parent model used as initial and boundary conditions and some model parameters. These different simulations provide the library of ocean states used by the data assimilation system.

The WMOP data assimilation system is based on a Local Multimodel Ensemble Optimal Interpolation (EnOI) scheme. In our approach, background error covariances (BECs) are estimated by sampling three long-run simulations of the WMOP system with different initial/boundary forcing (coming from CMEMS MED and GLOBAL models) and momentum diffusion

20   parameters.
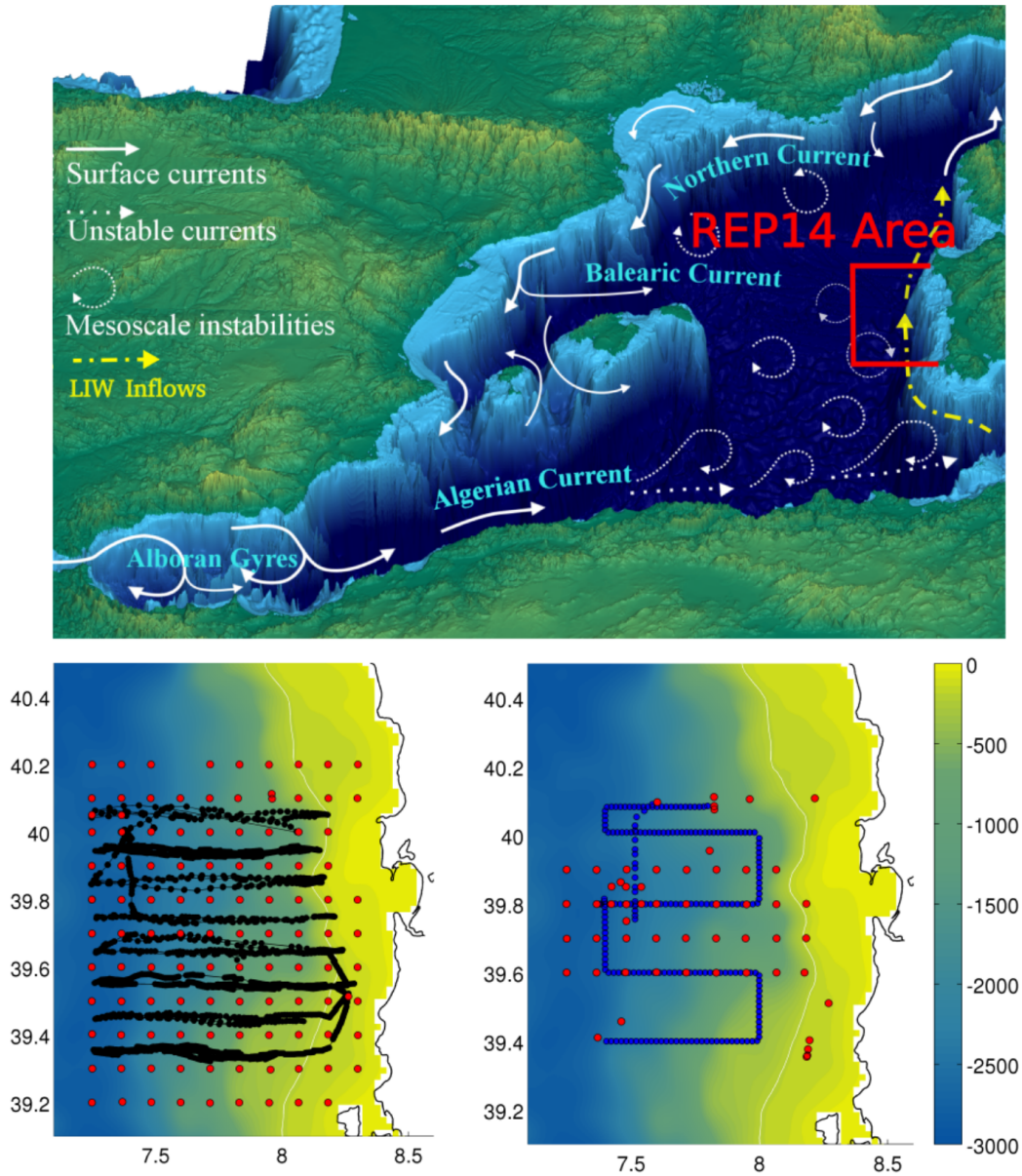
Figure 2. Top panel: Scheme of the principal circulation features in the western Mediterranean WMOP domain. Adapted from Escudier et al. (2016b); Millot (1999). The REP14-MED sea trial area is highlighted in red. The observations employed in this work are illustraed in the two bottom panels. CTDs from Leg 1 (red) and gliders (black) observations used for assimilation are shown in the bottom left panel. Independent CTDs (red) and Scanfish (blue) gathered during Leg 3 and used for the validation are shown in the bottom right pannel. The colorbar indicates depth and the white contour indicates the 200m isobath.

For each analysis, a 80-member ensemble is generated to calculate the BECs. Ensemble members are multivariate model fields sampled from the 3 simulations during the same season, with a time window of 90 days centered on the analysis date. The seasonal cycle is removed from the ensemble anomalies to discard the corresponding large scale correlations mainly affecting temperature. Following this procedure, the computed BECs reflect the spatial variability and anisotropy of the ocean mesoscale

5   circulation. They also represent the covariances between different model variables and depths in a dynamical consistent way. Moreover, a domain localization strategy is used to dampen the impact of remote observations. A 200km localization radius is used, determined by both the size of mesoscale structures and the approximate distance between two Argo platforms in the Western Mediterranean basin.

In the EnOI, as in any other sequential data assimilation scheme, a special care needs to be brought to the model initialization

10   after analysis updates (Oke et al., 2008). When restarting the simulation from an analysis field, the multivariate initial fields may be violating some physical constrains, such as mass conservation. The model response to balance this state may generate some spurious waves and introduce noise into the system. To minimize such effects, a nudging approach has been implemented after the analysis in the assimilation cycles produced in this work. In concrete terms, after an analysis is computed, the model is restarted 24 hours before the analysis date applying a strong nudging term to the temperature, salinity and sea level equations

15   towards analyzed values. The time scale associated with this nudging term is 1 day. The nudging procedure reduces the model correction, but guarantees updated multivariate fields closer to the model equation balances, which limits instabilities.

The assimilation system implemented here uses a 3-day cycle (figure 3), which was determined according to the time spent by the gliders to complete one zonal transect. All measurements collected during these 3 days are considered synoptic in the data assimilation process. Altimeter sea level anomalies (SLA), satellite sea surface temperature (SST) and Argo temperature and

20   salinity profiles are assimilated over the whole WMOP domain. For each analysis, a 5-day time window in the past is defined to select Argo observations. This window corresponds to the interval between two profiles provided by a single platform. This ensures that every model point is bounded by at least one Argo profile within the 200km localization radius in each analysis. Concerning altimetry, the past 72-hour CMEMS along-track reprocessed filtered sea level observations are considered for the analysis. The SST field is given by the daily L4 near real time GHRSST JPL-MUR satellite-derived interpolated product

25   (https://podaac.jpl.nasa.gov/dataset/JPL-L4UHfnd-GLOB-MUR). The last available field before analysis is considered. The original 1-km resolution data is smoothed and interpolated onto a 10km-resolution grid to limit the number of observations considered for each analysis. The selected resolution is considered to be sufficient to represent the main circulation features and mesoscale structures present in this SST product, permitting at the same time an affordable computational cost.

The glider profiles are considered as vertical. The corresponding observations are binned vertically and a single value is

30   given for each model grid cell. The representation error is the addition of a vertical and an horizontal components. For each vertical level, the observed variance in the vertical grid cell is used as an approximation of the vertical representation error. In addition, the horizontal representation error variance is assumed to be $(0.25°C)^2$ and 0.05^2 for temperature and salinity (in practical salinity scale) measurements, respectively. CTD observations are binned vertically in a similar way before assimilation, considering the representation error in an analogous way.
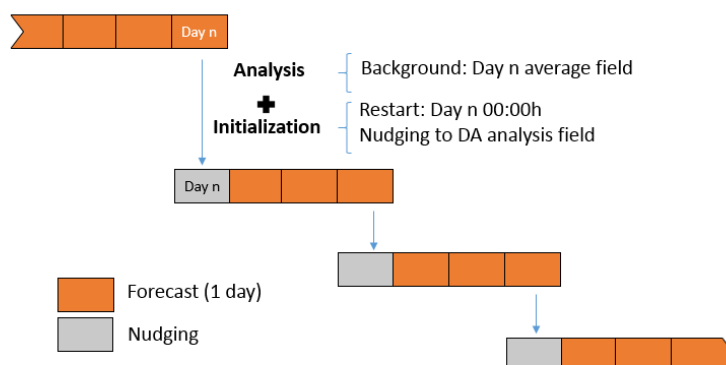
Figure 3. Scheme of the 3-day data assimilation cycles.

2.3 Experiments

In addition to the background simulation without any data assimilation (hereafter NO_ASSIM), seven simulations were produced spanning the period 1-24 June 2014, assimilating different sets of observations. The first simulation (GNR) assimilated generic observations from satellite along-track SLA, satellite SST and Argo temperature and salinity profiles. The second sim-

5 ulation (GNR_CTD) assimilated these generic observations plus all CTD temperature and salinity profiles collected during Leg 1. The five remaining simulations assimilated the generic observations plus glider temperature and salinity data from one to eight vehicles (GNR_1G, GNR_2G, GNR_3G, GNR_4G, GNR_8G), selected among the available platforms to optimally cover the area of interest. For example, GNR_1G considers the glider which travels in the center part of the domain, GNR_2G selects the two gliders which divide the study region in 3 areas of similar dimensions, etc... The different sets of vehicles

10 selected for these simulations are illustrated in Figure 4.

The whole timeline of the numerical experiments is described in Figure 5. A spinup period of 9 days was imposed for all these data-assimilative simulations, during which only the generic observations were assimilated. As explained previously, the 3-day assimilation cycle implemented in this study corresponds to the time spent by a glider to complete a zonal transect. In the case of the CTDs, as the duration of the data collection in Leg 1 was six days, the data was assimilated during two cycles.

15 After the last analysis, both Leg 3 CTDs and Scanfish temperature and salinity measurements are used as independent observations to evaluate the performance of the simulations.

3 Results

We evaluate in this Section the performance of the data assimilation following three successive steps. We first verify that the data from the different sources are properly ingested in the system over the whole modelling area both during the spinup

20 period and subsequent assimilation phase. Then, we examine the impact of the assimilation of the local and dense observations
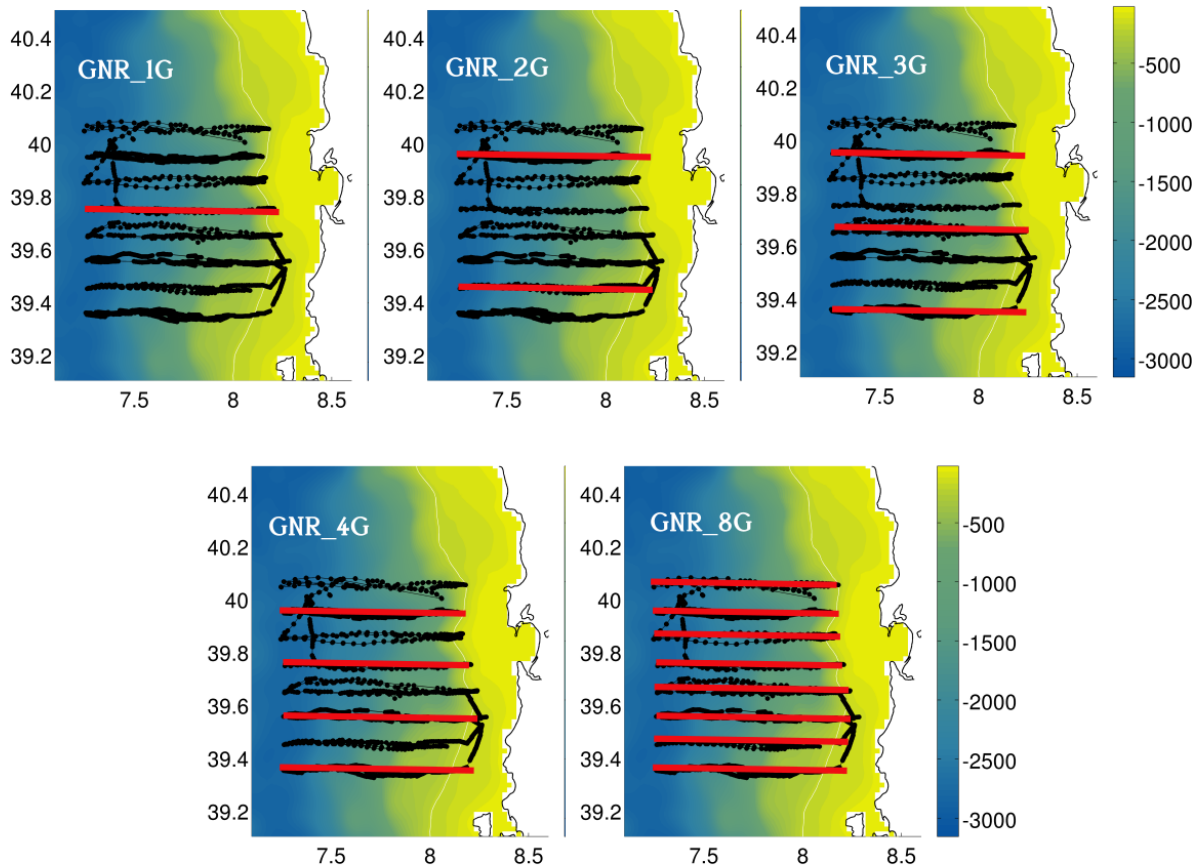
Figure 4. Illustration of the different sets of gliders selected in the different data-assimilative experiments. The position of all glider measurements are shown in black dots. The red zonal lines indicate the selected glider tracks in each of the experiments. The name of the corresponding simulation is specified in each panel.

datasets onto the temperature, salinity and density fields in the REP14-MED area. Finally, we assess the performance of the simulations against independent data from CTDs and Scanfish observations collected during Leg 3.

3.1 Data ingestion and performance over the whole modelling area

We first assess here the performance of the assimilation during the spinup period by analyzing the evolution of the Root-Mean-Square-Difference (RMSD) between the model simulations and satellite SLA, SST observations and Argo profiles. For each source and variable, the RMSD is calculated as expressed in equation 1 below, where o_i and m_i take the values of the observations and their model equivalents, respectively. n is the number of observations. To better highlight relative simulation improvements, the RMSD for each specific day is normalized by dividing the RMSD by that of the simulation without any
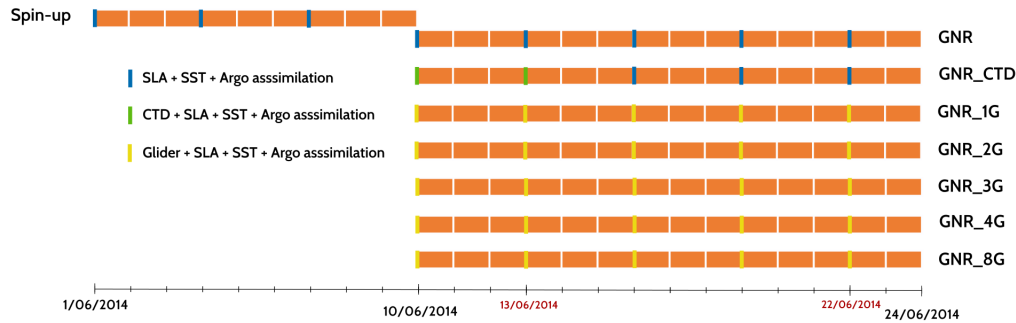
Figure 5. Timeline of the seven data-assimilative simulations. The analysis dates are highlighted in color, indicating the assimilated datasets.

data assimilation for that specific day. A reduction of the normalized RMSD indicates that the analyzed field is closer to the observations than the background field without assimilation.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n}(o_i - m_i)^2}{n}} \qquad\qquad RMSD_{normalized} = \frac{RMSD_{assim\_simulation}}{RMSD_{no\_assim}} \qquad\qquad (1)$$

The normalized RMSD is computed every day from 1 to 9 June. For the days including an analysis, the observations
5 assimilated during this analysis are used to compute the normalized RMSD. This includes the different assimilation windows (5 days for Argo, 3 days for SLA, one day for SST in particular). For the remaining days, we consider the observations that the system would have ingested if we had performed the analysis on that date, so considering similar time windows. Model equivalents to the observations are obtained through linear interpolation in space of the average daily model fields onto the position of the measurements.

10 The results are presented in figure 6. They show a satisfactory and continuous reduction of the RMSD for all the sources of data and variables, indicating a good system performance. The normalized RMSD is significantly reduced during the first analysis (between 20 and 60% depending on the analyzed variable), it then tends to slightly increase until the next analysis 3 days later, which reduces it again in most of the cases. In some occurrences, the RMSD continues decreasing during two days after the analysis. The overall persistence of the correction between two assimilation dates is especially remarkable. It reveals
15 the general proper performance of the assimilation system, which is able to recursively correct the multivariate fields without introducing spurious structures and instabilities which would significantly alter the system.

After this verification of the overall satisfactory performance of the data assimilation during the spin-up period in terms of RMSD, the same kind of assessment was performed for the six simulations assimilating either CTDs or glider observations during the field experiment. We only show here results from the GNR_CTD simulation (figure 7), since the behaviour is very
20 similar for the rest of the simulations. The system still properly reduces the normalized RMSD in terms of SST, SLA and T-S profiles at Argo locations, with similar reductions as that observed during the spinup period (from 20 to 60% of error reduction). An important aspect in this comparison is that the assimilation of high resolution T-S profiles data in the REP14-MED area
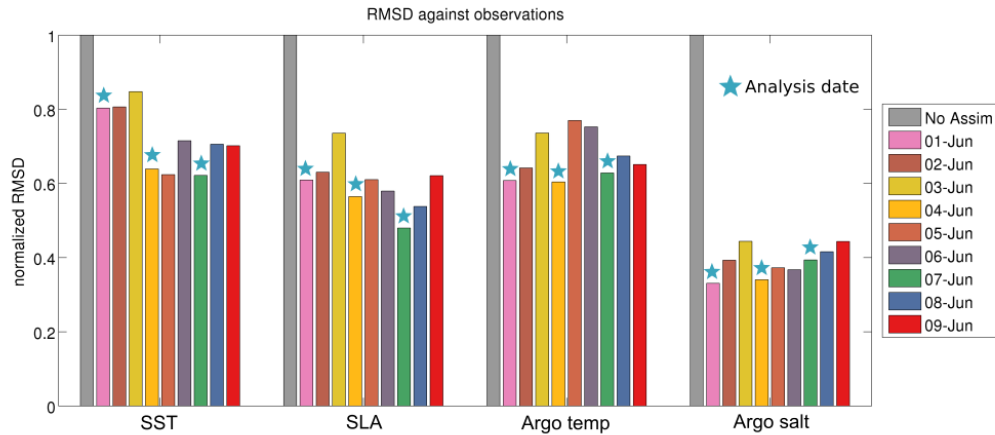
Figure 6. Evolution of the normalized RMSD against observations for the spinup simulation.

does not negatively affect the overall performance of the system over the whole modelling area. This could happen through the generation of spurious structures in the densely observed area which could then propagate over the domain. Moreover, the reduction of the normalized RMSD with respect to CTD observations shows that the local observations have also been properly ingested in the system.
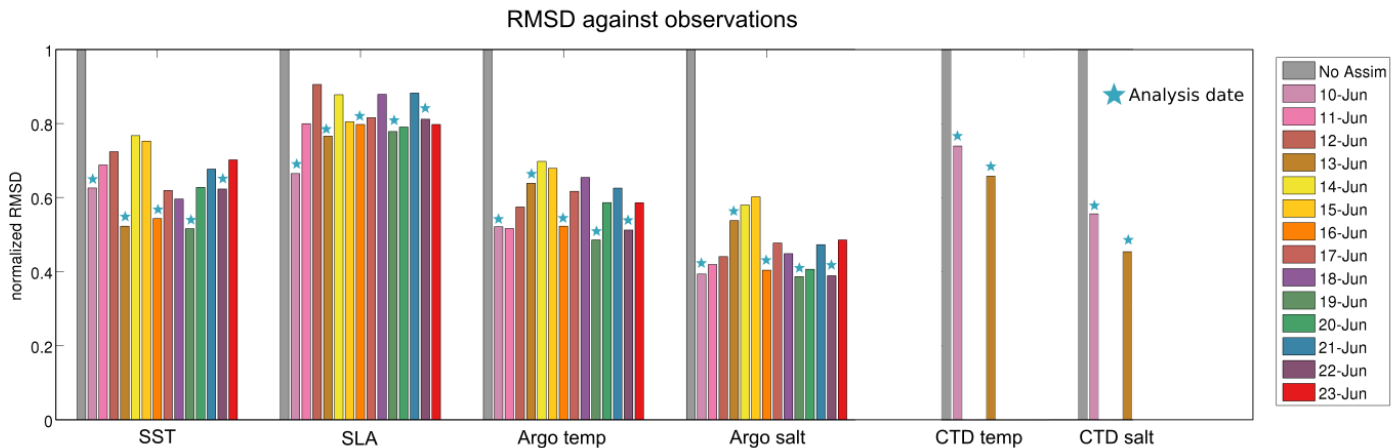
Figure 7. Evolution of the normalized RMSD against observations for the GNR_CTD simulation.

5 **3.2 Temperature, salinity and density fields in the REP14-MED area**

To complement these statistical diagnostics based on the normalized RMSD, we analyze here the temperature and salinity fields in the REP14-MED trial area on 13 June. This corresponds to the first day after the second analysis of CTD and glider

data assimilation cycles. At that time, either all CTD data from Leg 1, or one back-and-forth transect from the gliders have been introduced into the system. Figure 8 shows temperature, salinity and potential density daily average fields at 50 m depth for four of the simulations (NO_ASSIM, GNR, GNR_CTD and GNR_8G) on 13 June. The temperature and salinity data assimilated until that date are also represented as colored dots on the panels corresponding to GNR_CTD and GNR_8G.

5 Two different water masses are represented in the NO_ASSIM simulation. While the northern part of the domain is mostly occupied by a denser water mass with a salinity over 38, lower density waters characterized by their relative fresher and warmer characteristics are found in the southern part of the REP14-MED domain. The GNR simulation redistributes these water masses over the domain, representing patches of denser water in the central, south-western, north-western and northern coastal parts of the domain. These patches, with relatively warm and salty characteristics, are associated with cyclonic circulations.

10 The additional assimilation of dense local data from CTDs and gliders further modulates these patterns, producing smaller scale patches and filaments. Two main warm temperature anomalies are detected in both CTD and glider observations at 50m depth, associated with relatively slightly salty anomalies. The strongest one, located around 7.7E-39.3N, is somehow represented in both simulations GNR_CTD and GNR_8G, with a more pronounced signature in GNR_8G. Notice that this signature is not fully coincident with the observed location displayed in Figure 8 due to the evolution of the model from the

15 first analysis on 10 June to the time of the plot 3 days later. The second relatively warm temperature patch found around 7.5E-40N is less marked that the first one. Again, it is somehow better reproduced in GNR_8G than GNR_CTD. The observations, from both CTDs and gliders, are characterized by an energetic small scale variability, which translates into small scales and filamental structures in the model after data assimilation. Notice also the improvements in the relatively higher salinity along the coast after assimilation of the measurements from the CTDs and the gliders. The relatively high salinity patch around 39.5N

20 seen in the GNR simualtion is strongly attenuated in both GNR_CTD and GNR_8G. The density fields of GNR_8G exhibit two areas of lower potential density associated with these two anomalies, in good qualitative agrement with the observations even if th magnitude of the gradients is reduced compared to the measurements. These density anomalies are less clear in GNR_CTD. Both simulations show denser water on the north-eastern part of the domain and similar overall circulation patterns which significantly differ from NO_ASSIM and have also marked differences with GNR. A common property observed in both

25 simulations is the current flowing northeastwards in the central part of the sampled area and bifurcating near the coast, with one branch directed southwards and other northwards, giving also rise to cyclonic and anticyclonic eddies with dimensions around 30-40km.

3.3 Performance assessment using independent data during Leg 3

As a third step we analyze here the realism of the simulations during Leg 3 using independent observations which have not been

30 assimilated in the experiments. More specifically, we compared the model outputs on 22 June (after all assimilation cycles have been completed) with CTD and Scanfish temperature and salinity observations collected between 20 and 23 June. A qualitative analysis is first performed, based on the potential density fields reconstructed from both CTD and Scanfish observations at 50m depth. The DIVA software (Data Interpolating Variational analysis, (Barth et al., 2010)) and its web interface (http://ec.oceanbrowser.net/emodnet/diva.html) have been used to generate the interpolated density field. Figure 9 compares the
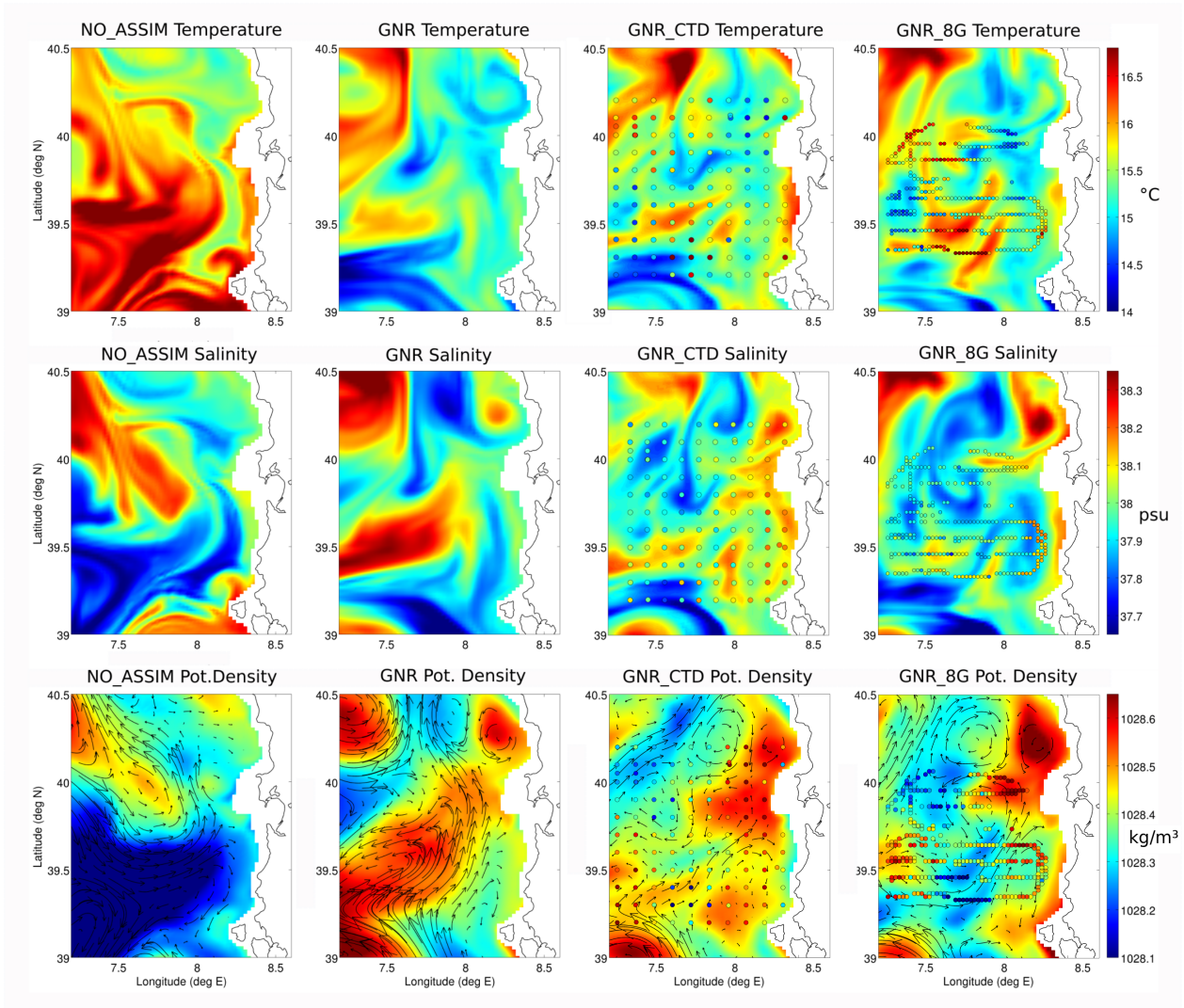
Figure 8. Model temperature (upper panels), salinity (middle panels) and potential density and currents (lower panels) at 50m depth on 13 June. From left to right: simulations NO_ASSIM, GNR, GNR_CTD and GNR_8G. The assimilated data are superimposed as colored dots in the temperature and salinity panels for the two simulations GNR_CTD and GNR_8G.

density fields at 50 meter depth from the different simulations with this density field derived from the observations. Notice that while the range of the colobar is the same for all the panels, the min and max are shifted by 0.2 kg/m^3 in the panels corresponding to the simulations due to a persistent bias in the model density fields. The CTD and Scanfish observations overplotted to the simulations fields use the same colorbar as in the reconstruction in the upper left panel.

5   The main features represented in the reconstructed density field derived from the observations include a marked negative density anomaly centered around 7.8E-39.4N with a spatial extension around 40km, a coastal fringe with relatively denser

waters, and a second patch of denser water between 39.5N and 40N on the Western side of the domain. These features were somehow already present during Leg 1 (see figure 8 and Section 3.2).

All the data-assimilative simulations represent the denser coastal fringe and the southward flow associated to it, yet with different characteristics. It extends offshore, associated with a cyclonic eddy, in GNR, GNR_CTD, GNR_1G and GNR_2G.

5   GNR_4G and GNR_8G qualitatively provide a more accurate shape of this coastal feature. In addition, these two simulations better represent the secondary relatively denser patch on the western side. Lower density anomalies south of 39.5N are also present in all the simulations. GNR, GNR_CTD and GNR_1G seem to qualitatively better match the reconstructed field by representing an anticyclonic eddy around a local density minimum, with an approximate 40km diameter. However, the exact shape of this anomaly and in particular its meridional extension, was not properly observed during Leg 3, which only provided

10 a single Scanfish zonal section at 39.4N across this anomaly. While it is represented as a close eddy in the reconstructed field due to the interpolation method, the more elongated shape in the meridional direction provided by GNR_4G and GNR_8G is also consistent with the Scanfish observations. Notice that the data-assimilative simulations all qualitatively improve the solution without data assimilation. Among them, GNR_4G and GNR_8G provide a particularly remarkable pattern agreement with the Scanfish observations.
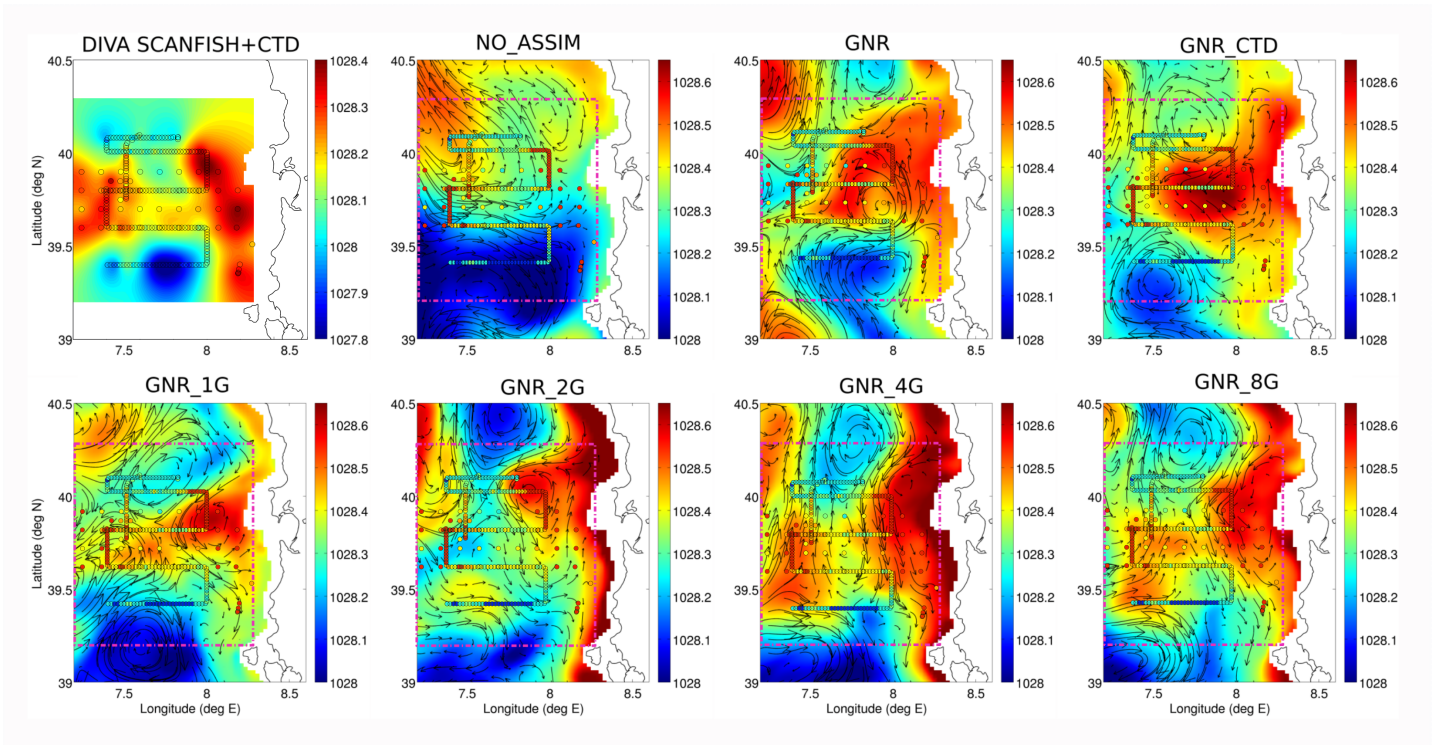
Figure 9. Top left panel: potential density field reconstruction from Scanfish and CTD data collected between 20 and 23 June (kg/m3). Remaining panels: potential density (kg/m3) and model currents at 50m depth on 22 June for the seven simulations NO_ASSIM, GNR, GNR_CTD, GNR_1G, GNR_2G, GNR_4G, GNR_8G and GNR_CTD.

To quantify the improvement, we now present the normalized RMSD, both considering the type of large scale observations which were assimilated over the whole domain, and the independent sea trial observations used in this Section. We computed the normalized RMSD for each of the seven data-assimilative simulations on 22 June and for the different sources of observation (figure 10). CTD and Scanfish observations between 20 and 23 June were considered synoptic for this purpose.
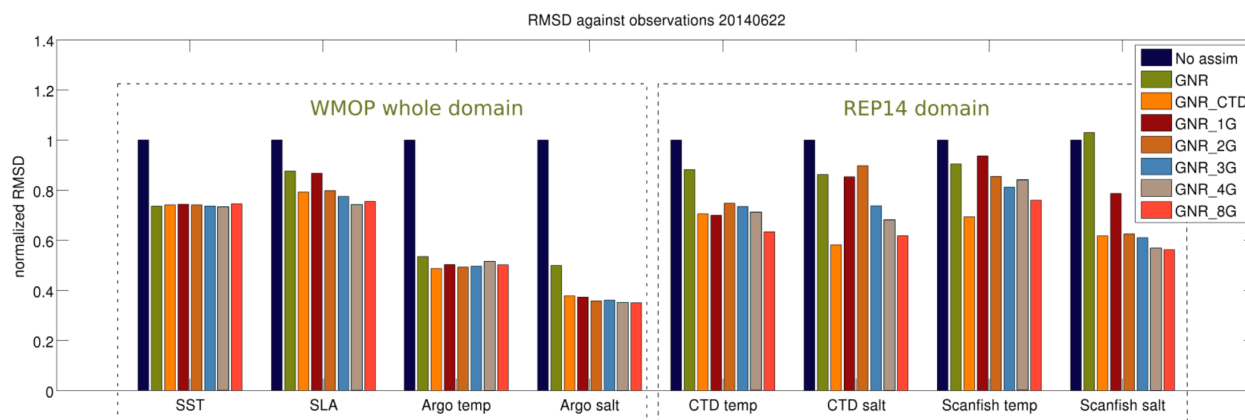
Figure 10. Normalized RMSD against observations on 22 June for the 7 numerical simulations. Dashed bounding boxes delimitate on the one side the observations assimilated over the whole domain, and on the other side the independent campaign observations within the REP14-MED domain.

5   As already described in Section 3.1, the generic assimilation (SLA along-track, SST and Argo) provides similar results over the whole domain to that obtained during the spinup period, when no dense profile data is assimilated in the REP14-MED domain. It reduces significantly the RMSD compared to the NO_ASSIM simulation. Moreover, it also allows to reduce by around 10% the RMSD against independent CTD observations both in temperature and salinity. While it improves the comparison with Scanfish temperature observations, it slightly degrades salinity comparisons.

10   The ingestion of high resolution local data from the REP14-MED campaign further reduces the RMSD with SST, SLA and Argo computed over the whole domain, with similar results when assimilating observations from CTDs and gliders (with the exception of the SLA which is not improved when a single glider is assimilated). In the REP14-MED domain, the assimilation of CTDs allows a reduction of the RMSD against independent observations between 30 and 40%, both in temperature and salinity, with respect to the simulation without any data assimilation. The assimilation of glider observations also reduces the

15   RMSD, with an overall enhanced performance as the number of platforms increases. The comparison with different platforms and variables provides slightly different rankings of the simulations. For instance, in this comparison, GNR_1G provides a similar performance as GNR_CTD against independent CTD temperature data, but a lower performance against CTD salinity and Scanfish measurements. The assimilation of data from 4 gliders improves the performance with respect to the assimilation of CTDs when comparing to Scanfish salinity data, but the performance is lower when comparing to the other sources. GNR_CTD

20   provides the best RMSD reduction when comparing to CTD salinity and Scanfish temperature, but it is GNR_8G which shows

the best performance when considering CTD temperature and Scanfish salinity data. These variations are probably due to the specific spatial sampling of the CTDs and Scanfish (see Figure 2) combined to the high spatial oceanic variability in the area.

An average RMSD reduction number is obtained here by computing the square root of the average normalized mean square difference over the four comparisons (CTD and Scanfish temperature and salinity) in the REP14-MED domain. These syn-

5 thetic average normalized RMSD scores are presented in Figure 11. This average RMSD gives scores of 39% and 40% of error reduction for the simulations GNR_CTD and GNR_8G, respectively, situating the CTD survey more performant than a sampling using 4 gliders, with a very close performance as that obtained with 8 gliders. Notice that the average normalized RMSD illustrates the progressive gain obtained when increasing the number of gliders considered in these experiments.
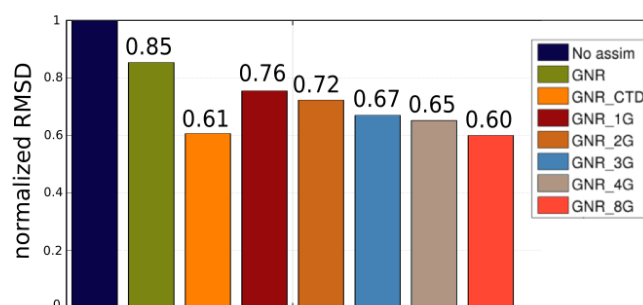
Figure 11. Average normalized RMSD against independent observations in the REP14-MED areas on 22 June for the 7 numerical simulations.

4 Discussion

10 The assimilation of observations is crucial to improve forecasts. Efficient regional forecasting systems should be able to efficiently combine high resolution local profile data and larger scale satellite observations over an extended modelling domain. The recursive Ensemble Optimal Interpolation scheme employed in this study is shown to be able to ingest both types of data and to systematically reduce the errors when compared to a control free-run simulation. Even if the EnOI is theoretically inferior to more advanced data assimilation schemes such as the ensemble Kalman filter or the 4DVar (see for instance Sakov and

15 Sandery (2015) and Oke et al. (2007) for comparisons of DA schemes), its numerical efficiency makes it a good compromise for operational and practical implementations with high resolution models.

The domain localization approach guarantees that the assimilation of dense profile observations from gliders and CTDs over a reduced area does not degrade the results over the whole modelling domain. Moreover, it allows to reduce the RMSD and to improve the representation of local water masses and the associated circulation in the reduced REP14-MED area which has

20 dimensions around 100km. The corrections introduced by the assimilation of CTD data during Leg 1 are found to remain in

time, providing a very positive and significant error reduction when comparing to independent measurements 10 days after the initial CTD data collection.

As a limitation, we notice that the oceanographic structures of small horizontal and vertical dimensions, which have a strong signature in the dense observation datasets, are only approximately represented in the temperature and salinity fields just after

5   assimilation, as shown in figure 8). This is the case for instance for the strong positive temperature anomaly around 7.7E-39.3N, which is the signature of an eddy with an horizontal diameter around 40 km and a vertical dimension around 50m. We attribute this limitation on the one hand to the smoothing effect of the background error covariances, which impacts both along the horizontal and vertical directions, and on the other hand to the nudging initialization procedure, which attenuates the model correction with the aim to provide more dynamically consistent fields. To illustrate the error covariances of our EnOI

10 implementation, figure 12 shows both the horizontal and vertical model ensemble correlations generated from the ensemble for the analysis on 22 June.
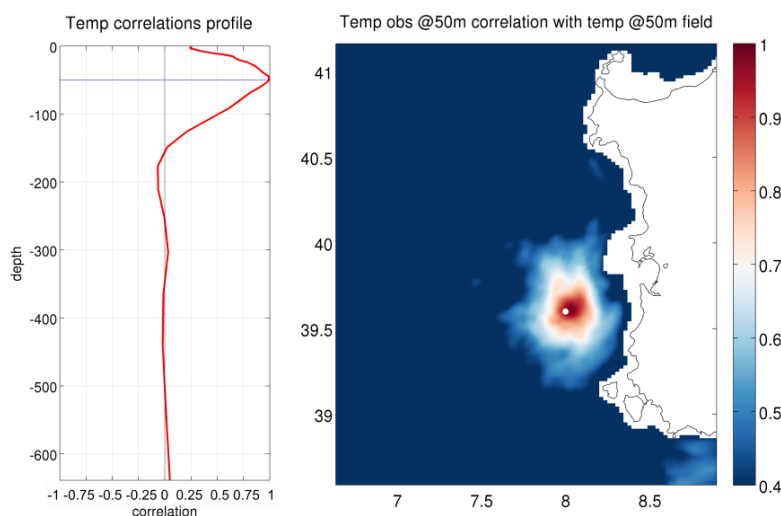
Figure 12. EnOI temperature ensemble correlations for a temperature observation within the REP14-MED domain at 50m depth (position indicated by the white dot on the right panel). Left: correlations along the vertical. Right: horizontal correlations at 50m depth.

A spatial smoothing takes place during the assimilation, affecting the area with significant correlations with the observed locations. The ensemble correlation distances are found to be around 100m in the vertical and 75km in the horizontal, being then larger than the smaller scale structures observed in the CTD and glider surveys. Two-steps assimilation strategies sep-

15 arating long- and short-distance correlation scales might allow to improve the representation of these finer patterns in more sophisticated data assimilation systems (e.g. Li et al. (2015)).

The second factor explaining this limitation is related to the nudging initialization strategy, which has the advantage of limiting undesired model shocks after the analysis, but also attenuates the corrections and therefore the agreement with obser-

vations. The simple nudging technique used in this work is easy to implement and cost-effective. It could be improved in the future by considering more advanced approaches (e.g. (Sandery et al., 2011)).

In spite of these limitations, the EnOI scheme implemeted in this study is shown to be able to properly ingest the multi-scale observations, which leads to improved representations of the mesoscale structures in the REP14-MED area and enhanced forecasting skills persisting several cycles after the assimilation of the dense observations.

While CTDs allow a relatively fast comprehensive description of a specific study area, gliders provide a slower sampling but also allow a repetition of specific monitoring tracks over a longer period. In this study, the CTD initialization survey results in a similar forecast performance after data assimilation as an 8-glider continuous monitoring of the area flying along predefined paths with regular spacing. Let's mention that while these platforms are considered autonomous, their operation still implies a very significant effort in terms of platform deployment, recovery, piloting and maintenance. The models could also highly benefit from the near real-time controllability of gliders, allowing to continuously adjust their path along optimal routes in the study area. In this framework, efficient adaptive sampling procedures should theoretically allow to use a reduced number of gliders to reach the same level of performance ((Lermusiaux, 2007)). The definition of optimal collective behaviours based for instance on glider fleet coordination or cooperation (e.g. (Alvarez and Mourre, 2014)) constitutes an interesting field of research in that direction.

5 Conclusions

We presented in this work the results of several simulations assimilating different multi-platform observations in the context of the REP14-MED sea trial carried out in June 2014 off the West coast of Sardinia. The experiments were designed to assimilate intensive campaign data from CTDs and gliders, along with satellite SST and SLA, as well as Argo profile observations over the whole model domain covering the Western Mediterranean Sea from Gibraltar to the Sardinia Channel. The objective was to explore the performance of different sampling strategies based on either a dense CTD initialization survey or a glider fleet sampling, in improving model forecasting capabilities in a specific area.

The data assimilation system was shown to perform correctly. The Local Multimodel EnOI scheme, following 3-day recursive cycles with a 1-day nudging initialization phase after analysis, allows to properly ingest both large scale data over the whole Western Mediterranean domain and high density temperature and salinity profiles collected during the sampling experiment over a limited area. In spite of the limitations associated with the smoothing effect of ensemble covariances, which do not allow to exactly represent the smaller scale features present in the observations, this system enables a significant improvement of the forecasting skill of the model with respect to the simulation without assimilation, and that assimilating only satellite and Argo data. Its reduced cost makes it a good option for operational implementations.

While the assimilation of generic observations from SST, SLA and Argo leads to an average error reduction of 15% when comparing to independent measurements collected during the third Leg of the sea trial in the REP14-MED area, the assimilation of glider and CTD data allow an additional significant improvement. Gliders, which provide a continuous sampling of the area along regularly spaced zonal tracks, allow to reduce the forecast error as the number of platforms increases. The consideration

of one glider leads to a 24% average error reduction with respect to the simulation without assimilation. This percentage increases to 28%, 33%, 35% and finally 40% for the 2-glider, 3-glider, 4-glider and 8-glider fleet configurations, respectively. Incrementing the number of gliders results in a better representation of the ocean state captured by observations, with a most accurate representation of the mesoscale structures and associated circulation.

5   The assimilation of the observations from the dense initialization survey based on 10-km spaced CTD stations leads to an average error reduction of 39%: it outperforms the 4-glider configuration and provides very similar results in terms of RMSD as the 8-glider fleet configuration. The 10km spacing offered by both sampling strategies is essential here to improve the representation of the mesoscale variability in the study area. In view of these results, gliders certainly provide a very interesting alternative to traditional CTD surveys used to initialize high-resolution regional ocean models, provided that a fleet

10 of vehicles can be deployed at sea. Moreover, an increased performance can certainly still be expected by optimizing the regular track sampling carried out in this experiment through adaptive sampling procedures.

Acknowledgements. The authors especially thank Reiner Onken for leading the experiment. They also acknowledge all the partners, scientists and technicians having participated to the REP14-MED sea trial, allowing the collection and processing of this very valuable dataset. SOCIB data assimilation system developments were partly supported by the Medclic project funded by La Caixa foundation and the JERICO-NEXT

15 European project.

References

Alvarez, A. and Mourre, B.: Cooperation or coordination of underwater glider networks? An assessment from observing system simulation experiments in the Ligurian Sea, Journal of Atmospheric and Oceanic Technology, 31, 2268–2277, https://doi.org/10.1175/JTECH-D-13-00214.1, 2014.

5 Barth, A., Alvera-Azcárate, A., Troupin, C., Ouberdous, M., and Beckers, J. .: A web interface for griding arbitrarily distributed in situ data based on Data-Interpolating Variational Analysis (DIVA), Advances in Geosciences, 28, 29, https://search.proquest.com/docview/834425514?accountid=10336, copyright - European Geosciences Union 2010; Última actualización - 2011-01-11, 2010.

Escudier, R., Mourre, B., Juza, M., and Tintoré, J.: Journal of Geophysical Research : Oceans, Journal of Geophysical Research: Oceans, pp. 914–940, https://doi.org/10.1002/2016JC012121.Received, 2016a.

10 Escudier, R., Renault, L., and Pascual, A.: Eddy properties in the Western Mediterranean Sea from satellite altimetry and a numerical simulation, Journal of Geophysical Research: Oceans, pp. 1–17, https://doi.org/10.1002/2015JC011371.Received, 2016b.

Evensen, G.: The Ensemble Kalman Filter: Theoretical formulation and practical implementation, Ocean Dynamics, 53, 343–367, https://doi.org/10.1007/s10236-003-0036-9, 2003.

Garau, B., Bonet, M., Alvarez, A., Ruiz, S., and Pascual, A.: Path planning for autonomous underwater vehicles in realistic oceanic current
15 fields: application to gliders in the western mediterranean sea, VI, 5–22, 2009.

Heslop, E. E., Ruiz, S., Allen, J., López-jurado, J. L., Renault, L., and Tintoré, J.: Autonomous underwater gliders monitoring variability at "choke points" in our ocean system: A case study in the Western Mediterranean Sea, 39, 1–6, https://doi.org/10.1029/2012GL053717, 2012.

Juza, M., Mourre, B., Renault, L., Gómara, S., Sebastian, K., López, S. L., Borrueco, B. F., Beltran, J. P., Troupin, C., Tomás, M. T., Heslop,
20 E., Casas, B., and Tintoré, J.: Operational SOCIB forecasting system and multi-platform validation in the Western Mediterranean, Journal of Operational Oceanography, p. 9231, https://doi.org/10.1002/2013JC009231., 2016.

Knoll, M., Borrione, I., Fiekas, H.-v., Funk, A., Hemming, M. P., Kaiser, J., Onken, R., Queste, B., Russo, A., and Geesthacht, H.-z.: Hydrography and circulation west of Sardinia in June 2014, pp. 889–904, 2017.

Lermusiaux, P. F.: Adaptive modeling, adaptive data assimilation and adaptive sampling, Physica D: Nonlinear Phenomena, 230, 172–196,
25 https://doi.org/10.1016/j.physd.2007.02.014, 2007.

Li, Z., McWilliams, J. C., Ide, K., and Farrara, J. D.: Coastal ocean data assimilation using a multi-scale three-dimensional variational scheme, Ocean Dynamics, 65, 1001–1015, https://doi.org/10.1007/s10236-015-0850-x, 2015.

Liblik, T., Karstensen, J., Testor, P., Alenius, P., Hayes, D., Ruiz, S., Heywood, K. J., Pouliquen, S., Mortier, L., and Mauri, E.: Potential for an underwater glider component as part of the Global Ocean Observing System, Methods in Oceanography, 17, 50–82,
30 https://doi.org/10.1016/j.mio.2016.05.001, http://dx.doi.org/10.1016/j.mio.2016.05.001, 2016.

Melet, A., Verron, J., and Brankart, J.: Potential outcomes of glider data assimilation in the Solomon Sea: Control of the water mass properties and parameter estimation, Journal of Marine Systems, 94, 232–246, https://doi.org/10.1016/j.jmarsys.2011.12.003, http://dx.doi.org/10.1016/j.jmarsys.2011.12.003, 2012.

Millot, C.: Circulation in the Western Mediterranean Sea, Journal of Marine Systems, 20, 423–442, https://doi.org/10.1016/S0924-
35 7963(98)00078-5, 1999.

Mourre, B. and Alvarez, A.: Benefit assessment of glider adaptive sampling in the Ligurian Sea, Deep-Sea Research Part I: Oceanographic Research Papers, 68, 68–78, https://doi.org/10.1016/j.dsr.2012.05.010, http://dx.doi.org/10.1016/j.dsr.2012.05.010, 2012.

Mourre, B. and Chiggiato, J.: A comparison of the performance of the 3-D super-ensemble and an ensemble Kalman filter for short-range regional ocean prediction, Tellus, Series A: Dynamic Meteorology and Oceanography, 66, 21 640, https://doi.org/10.3402/tellusa.v66.21640, 2014.

Oke, P. R., Sakov, P., and Corney, S. P.: Impacts of localisation in the EnKF and EnOI: Experiments with a small model, Ocean Dynamics, 57, 32–45, https://doi.org/10.1007/s10236-006-0088-8, 2007.

Oke, P. R., Brassington, G. B., Griffin, D. A., and Schiller, A.: The Bluelink ocean data assimilation system (BODAS), 21, 46–70, https://doi.org/10.1016/j.ocemod.2007.11.002, 2008.

Olita, A., Ribotti, A., Sorgente, R., Fazioli, L., and Perilli, A.: SLA–chlorophyll-a variability and covariability in the Algero-Provençal Basin (1997–2007) through combined use of EOF and wavelet analysis of satellite data, pp. 89–102, https://doi.org/10.1007/s10236-010-0344-9, 2011.

Onken, R., Fiekas, H.-v., Beguery, L., Borrione, I., Funk, A., Hemming, M., Heywood, K. J., Kaiser, J., Knoll, M., Poulain, P.-m., Queste, B., Russo, A., Shitashima, K., and Siderius, M.: High-Resolution Observations in the Western Mediterranean Sea: The REP14-MED Experiment, Ocean Science, https://doi.org/10.5194/os-2016-82, 2018.

Pan, C., Zheng, L., Weisberg, R. H., Liu, Y., and Lembke, C. E.: Comparisons of different ensemble schemes for glider data assimilation on West Florida Shelf, OCEAN MODELLING, 81, 13–24, https://doi.org/10.1016/j.ocemod.2014.06.005, http://dx.doi.org/10.1016/j.ocemod.2014.06.005, 2014.

Pascual, A., Ruiz, S., Olita, A., Troupin, C., Claret, M., Casas, B., Mourre, B., Poulain, P.-m., Tovar-sanchez, A., Capet, A., Mason, E., Allen, J. T., Mahadevan, A., and Oceanographic, W. H.: A multiplatform experiment to unravel meso- and submesoscale processes in an intense front (AlborEx), https://doi.org/10.3389/fmars.2017.00039, 2017.

Robinson, R. A., Leslie, G. W., Theocharis, A., and Lascaratos, A.: Mediterranean Sea Circulation, Ocean Currents, pp. 1–19, https://doi.org/10.1006/rwos.2001.0376, 2001.

Rudnick, D. L.: Ocean Research Enabled by Underwater Gliders, Annual Review of Marine Science, 8, 519–541, https://doi.org/10.1146/annurev-marine-122414-033913, http://www.annualreviews.org/doi/10.1146/annurev-marine-122414-033913, 2016.

Ruiz, S., Garau, B., Martínez-ledesma, M., Casas, B., Pascual, A., Vizoso, G., Bouffard, J., Heslop, E., Alvarez, A., Testor, P., and Tintoré, J.: New technologies for marine research: five years of glider activities at IMEDEA; Nuevas tecnologías para la investigación marina: 5 años de actividades de gliders en el IMEDEA, https://doi.org/10.3989/scimar.03622.19L, 2012.

Sakov, P. and Sandery, P. A.: Comparison of EnOI and EnKF regional ocean reanalysis systems, Ocean Modelling, 89, 45–60, https://doi.org/10.1016/j.ocemod.2015.02.003, 2015.

Sandery, P. A., Brassington, G. B., and Freeman, J.: Adaptive nonlinear dynamical initialization, Journal of Geophysical Research, 116, C01 021, https://doi.org/10.1029/2010JC006260, http://doi.wiley.com/10.1029/2010JC006260, 2011.

Santinelli, C., Ribotti, A., Sorgente, R., Gasparini, G. P., Nannicini, L., Vignudelli, S., and Seritti, A.: Coastal dynamics and dissolved organic carbon in the western Sardinian shelf (Western Mediterranean), Journal of Marine Systems, 74, 167–188, https://doi.org/10.1016/j.jmarsys.2007.12.005, http://dx.doi.org/10.1016/j.jmarsys.2007.12.005, 2008.

Shchepetkin, A. F. and McWilliams, J. C.: The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model, Ocean Modelling, 9, 347–404, https://doi.org/10.1016/j.ocemod.2004.08.002, 2005.

Shulman, I., Rowley, C., Anderson, S., Derada, S., Kindle, J., Martin, P., Doyle, J., Cummings, J., Ramp, S., Chavez, F., Fratantoni, D., and Davis, R.: Impact of glider data assimilation on the Monterey Bay model, Deep-Sea Reasearch II, 56, 188–198, https://doi.org/10.1016/j.dsr2.2008.08.003, 2008.

Simoncelli, Simona, F. C. P. N. G. A. D. M. O. P. . D. S.: Mediterranean Sea physical reanalysis (MEDREA 1987-2015) (Version 1)[Data set]. Copernicus Monitoring Environment Marine Service (CMEMS), https://doi.org/https://doi.org/10.25423/medsea_reanalysis_phys_006_004, 2017.

Testor, P., Be, K., Mortier, L., Sea, T., and Basin, A.: Modeling the deep eddy field in the southwestern Mediterranean: The life cycle of Sardinian eddies, 32, 7–10, https://doi.org/10.1029/2004GL022283, 2005.

Testor, P., Meyers, G., Pattiaratchi, C., Bachmayer, R., Hayes, D., Pouliquen, S., de la Villeon, L. P., Carval, T., Ganachaud, A., Gourdeau, L., Mortier, L., Claustre, H., Taillandier, V., Lherminier, P., Terre, T., Visbeck, M., Karstensen, J., Krahmann, G., Alvarez, A., Rixen, M., Poulain, P.-M., Osterhus, S., Tintore, J., Ruiz, S., and Garau, B.: Gliders as a Component of Future Observing Systems, Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society, pp. 961–978, https://doi.org/10.5270/OceanObs09.cwp.89, http://www.oceanobs09.net/proceedings/cwp/cwp89, 2010.