

A comparison of the performance of the 3-D super-ensemble and an ensemble Kalman filter for short-range regional ocean prediction

By BAPTISTE MOURRE^{1*†} and JACOPO CHIGGIATO², ¹*Centre for Maritime Research and Experimentation, Viale San Bartolomeo 400, 19126 La Spezia, Italy;* ²*Consiglio Nazionale delle Ricerche–Istituto di Scienze Marine (CNR-ISMAR), Arsenale – Tesa 104 Castello 2737/F, 30122 Venezia, Italy*

(Manuscript received 7 June 2013; in final form 9 December 2013)

ABSTRACT

This study compares the ability of two approaches integrating models and data to forecast the Ligurian Sea regional oceanographic conditions in the short-term range (0–72 hours) when constrained by a common observation dataset. The post-processing 3-D super-ensemble (3DSE) algorithm, which uses observations to optimally combine multi-model forecasts into a single prediction of the oceanic variable, is first considered. The 3DSE predictive skills are compared to those of the Regional Ocean Modeling System model in which observations are assimilated through a more conventional ensemble Kalman filter (EnKF) approach. Assimilated measurements include sea surface temperature maps, and temperature and salinity subsurface observations from a fleet of five underwater gliders. Retrospective analyses are carried out to produce daily predictions during the 11-d period of the REP10 sea trial experiment. The forecast skill evaluation based on a distributed multi-sensor validation dataset indicates an overall superior performance of the EnKF, both at the surface and at depth. While the 3DSE and EnKF perform comparably well in the area spanned by the incorporated measurements, the 3DSE accuracy is found to rapidly decrease outside this area. In particular, the univariate formulation of the method combined with the absence of regular surface salinity measurements produces large errors in the 3DSE salinity forecast. On the contrary, the EnKF leads to more homogeneous forecast errors over the modelling domain for both temperature and salinity. The EnKF is found to consistently improve the predictions with respect to the control solution without assimilation and to be positively skilled when compared to the climatological estimate. For typical regional oceanographic applications with scarce subsurface observations, the lack of physical spatial and multivariate error covariances applicable to the individual model weights in the 3DSE formulation constitutes a major limitation for the performance of this multi-model-data fusion approach compared to conventional advanced data assimilation strategies.

Keywords: operational oceanography, regional ocean prediction, data assimilation, ensemble Kalman filter, 3-D super-ensemble, underwater gliders, Ligurian Sea

1. Introduction

Predicting the evolution of physical ocean parameters has become a routine activity in operational ocean forecasting centres worldwide. Actively developed from the early 2000s, the global and basin-scale operational ocean predictions at days-to-weeks time scales (Dombrowsky et al., 2009) have now become mature enough to provide initial

and boundary conditions to operationally force nested regional ocean circulation models with an increased resolution. Regional ocean predictions at hours-to-days time scales are now operationally produced using this model nesting strategy (e.g. Zodiatis et al., 2008; Chao et al., 2009; Marta-Almeida et al., 2011). Data assimilation, which aims at optimally combining numerical dynamical ocean models with existing in-situ and remotely sensed observations, constitutes an essential component of these systems since it helps to recursively improve the initial conditions used for the prediction phases. Different data assimilation approaches may be adopted,

*Corresponding author.

email: bmourre@socib.es

†Now at: SOCIB, Parc Bit, Naorte A, 2^o3, 07121 Palma de Mallorca, Spain.

conventionally either from the sequential or variational families. Ensembles of model solutions are regularly used to approximate the model error covariances (e.g. Brasseur et al., 2005; Oke et al. 2008; Sakov et al., 2012).

Inspired from atmospheric applications (e.g. Krishnamurti et al., 1999; Doblas-Reyes et al., 2005; Raftery et al., 2005), multi-model ocean forecasting approaches have recently raised a particular interest due to the availability of several and sometimes inconsistent operational ocean predictions over a given region (Logutov and Robinson, 2005; Rixen et al., 2009; Lenartz et al., 2010; Mourre et al., 2012; Wang et al., 2013). These approaches generate a single consensual forecast by integrating all available information from the models and observations, which might then advantageously be used in decision-making processes. Proposed by Lenartz et al. (2010), the 3-D super-ensemble (3DSE) provides a post-processing methodology to produce a single optimal prediction when multiple forecasts and observations are available, without having to run an additional numerical ocean model. In this method, the spatially varying weights applied to the individual models are adaptively optimised based on the mismatch between observations and the weighted linear combination of the models during a recent learning period. While the definition of spatially varying model weights intends to cope with the spatial variability of individual model skills, spatial covariance functions are used to smooth and extrapolate their values outside the geographical area that bounds the observations. Using the operational outputs from three high-resolution models, the 3DSE was shown to lead to improved short-term ocean temperature predictions in the Ligurian Sea in 2008 in terms of the statistical measures of observation-model differences (Lenartz et al., 2010). Using an independent dataset collected 2 yr later in the same area, Mourre et al. (2012) (i) confirmed this potential, yet with a slightly reduced performance compared to Lenartz et al. (2010), and (ii) validated the associated uncertainty prediction. Although similarly based on a Bayesian formulation, the data integration in the 3DSE differs from conventional data assimilation in that observations are used to optimise the spatially variable weights of the different models instead of the physical variables themselves. The 3DSE analysis uses distance-based analytical covariance functions specified in the space of model weights to spatially propagate the information over the modelling domain. The lack of multivariate model weight error covariances makes the present formulation of the method univariate, which means that separate simulations have to be run for the prediction of the different variables. Mourre and Alvarez (2012) reported the outcomes of a glider adaptive sampling experiment driven by 3DSE uncertainty predictions in the Ligurian Sea. While the local ocean temperature forecast accuracy was successfully improved thanks to the

glider measurements, the 3DSE ocean velocity predictions exhibited significant errors, which were in part attributed to the univariate formulation of the method.

The present article provides a more comprehensive evaluation of the performance of the 3DSE approach for short-term ocean forecasting by comparing the corresponding predictions to those obtained through a more standard modelling and data assimilation method when constrained by a common dataset including observations from a fleet of gliders. A stochastic multivariate ensemble Kalman Filter approach (EnKF, Evensen, 2003) is used for this comparison. The EnKF is implemented with a limited-area configuration of the Regional Ocean Modeling System (ROMS, Haidvogel et al., 2008) nested in the Mediterranean Forecasting System (MFS, Pinardi and Coppini, 2010).

The region under study is the Ligurian Sea, where the 3DSE predictions were evaluated favourably with reference to the individual model forecasts and their average (Lenartz et al. 2010; Mourre et al. 2012). The extensive set of observations collected in summer 2010 during the Recognized Environmental Picture 2010 (REP10) oceanographic experiment is used to carry out this skill assessment. A common dataset is used to constrain both methods, which are implemented to retrospectively produce daily 72-hour ocean predictions from 21 to 31 August 2010. This dataset includes temperature and salinity observations from five gliders as well as high-resolution optimally interpolated sea surface temperature (OISST) from satellite data. It constitutes a typical modern ocean sampling scenario with regular temperature observations at the surface and sparse but high-resolution sections of temperature and salinity from robotic platforms in the ocean interior. The 3DSE and EnKF temperature and salinity forecast skills are evaluated against measurements from (i) a Conductivity, Temperature, Depth (CTD) profiler housed in a towed, undulating *Scanfish* vehicle, (ii) an oceanographic mooring, (iii) ship surface CTD transects and (iv) 65 vertical CTD profiles distributed over the modelling area.

The paper is organised as follows. Section 2 presents the experimental design including the description of the sea trial experiment, the prediction systems and the assimilated observations. Section 3 presents the performance of the systems in predicting ocean temperature and salinity. The results are finally discussed and summarised in Sections 4 and 5.

2. Experimental design

2.1. Ligurian Sea REP10 sea trial experiment

The REP10 experiment, with the objective of improving the rapid characterisation of the marine environment, was carried out in the Ligurian Sea from 19 August to

3 September 2010. The NATO Research Vessel (NRV) Alliance was used to collect ocean temperature and salinity measurements through a continuous shipborne surface CTD, a CTD housed on-board a towed undulating *Scanfish* vehicle and CTD profiles at fixed stations. A fleet of four shallow water *Slocum* underwater gliders equipped with CTDs was deployed from 20 to 29 August, providing continuous temperature and salinity sampling from the surface to 200 m depth. In addition, a deep-diving *Spray* glider was also deployed between 21 and 25 August to provide oceanic conductivity and temperature measurements down to 420 m depth. Concurrently, the permanent Ligurian Sea Ocean Data Acquisition System (ODAS Italia 1 mooring) provided temperature time series at the fixed location $43.84^{\circ}\text{N}/9.11^{\circ}\text{E}$ for five discrete levels from the surface to 36 m depth (specifically at 1, 6, 12, 20 and 36 m depth).

Figure 1 illustrates the satellite sea surface temperature measured in the Ligurian Sea by the Advanced Very High Resolution Radiometer (AVHRR – NOAA TIROS-N) on 21 August 2010, together with the position of the ODAS mooring and the trajectories of the five gliders during the whole duration of the experiment. The gliders travelled in the frontal region separating the warm coastal waters from the colder waters present in the central Ligurian Sea. This region is also characterised by the presence of mesoscale and submesoscale variability associated with this thermal front. The main flow enters the domain from the southern boundary along the western Corsica shelf and exits in the northern part of the western boundary after being cyclonically deflected due to the shelf and coastline geometry.

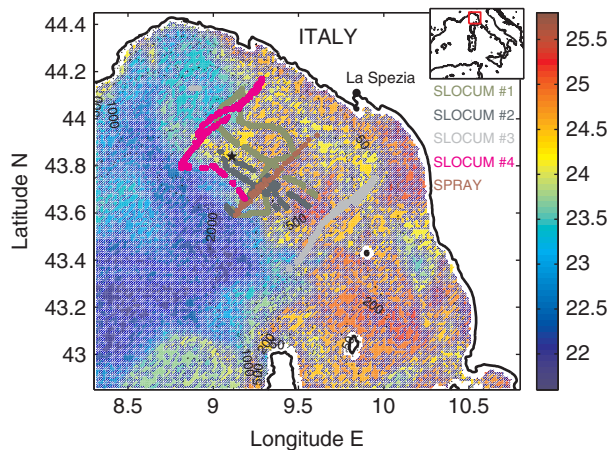


Fig. 1. Ligurian Sea surface temperature ($^{\circ}\text{C}$) on 21 August 08:34 (from AVHRR satellite) and trajectories of the five assimilated gliders between 21 and 31 August. The position of the ODAS mooring ($43.84^{\circ}\text{N } 9.11^{\circ}\text{E}$) is indicated by the black star.

2.2. Prediction systems

2.2.1. EnKF implemented with ROMS. The ocean circulation model used in this study is a limited-area configuration of the ROMS (Shchepetkin and McWilliams, 2005; Haidvogel et al., 2008). ROMS is a primitive-equation, finite-difference, hydrostatic and free-surface model using generalised terrain-following s -coordinates. The regional modelling domain covers the full Ligurian Sea, with two open boundaries located on the western (8°E) and southern (42.5°N) sides. The horizontal resolution is 1.8 km. The vertical grid is composed of 32 vertical s -levels non-linearly stretched to resolve the surface boundary layer. The momentum, heat and freshwater fluxes at the model surface are computed using the outputs from the 7-km resolution atmospheric model COSMO-ME of the Italian Air Force National Meteorological Center (Centro Nazionale per la Meteorologia e Climatologia Aeronautica CNMCA, Bonavita and Torrisi, 2005). The algorithm proposed by Marchesiello et al. (2001) combining radiation and nudging is implemented to calculate the baroclinic flows through the lateral open boundaries, using external values from the forecasts of the large scale (MFS, Pinardi and Coppini, 2010). The model simulates climatological runoffs from the rivers Arno, Magra and Serchio. The model was initialised on 1 May 2010 using an analysis field from the MFS model and run in forecast mode without assimilation until the beginning of the REP10 experiment on 21 August 2010. Further details concerning the model configuration can be found in Alvarez et al. (2013).

The EnKF (Evensen, 2003) is an advanced sequential data assimilation scheme using an ensemble of perturbed model simulations to approximate the model error covariances and their evolution in time. In this approach, the ensemble mean is considered as the best estimate of the state of the system. The EnKF has been successfully employed to assimilate different types of observations in the atmosphere and in the ocean (e.g. Dowell et al. 2004; Keppenne et al. 2005; Leeuwenburgh 2007) and is presently being used in several weather and ocean operational prediction centres (e.g. Houtekamer et al., 2009; Bonavita et al., 2010; Sakov et al., 2012). An asynchronous formulation (Sakov et al., 2010) is used here to allow the assimilation of observations from a time different from that of the analysis. A 24-hour assimilation window centred every day at 12:00 is used to generate the updated ensemble initial conditions for this central time. The daily 72-hour forecast starting at 00:00 is then preceded by a 12-hour ensemble spin up, so as to limit the impact on the prediction by potential spurious numerical transients related to the model restart after the analysis.

Given the vector of observations y collected between t_{i-1} and t_i and an ensemble of model forecasts run during this

period, the EnKF produces an analysis ensemble for a specified time t_j between t_{i-1} and t_i . The ensemble mean augmented state vector $\bar{\mathbf{x}}$, containing the model state variables at time t_j and the model values at the position and time of the observations collected between t_{i-1} and t_i , is updated according to the following equation:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \tilde{\mathbf{K}}(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f) \quad (1)$$

where the overbar denotes the ensemble mean and the superscripts f and a represent the forecast and analysis states, respectively. $\mathbf{H}$ is the linear observation operator projecting the model state onto the observation space and $\tilde{\mathbf{K}}$ is the Kalman gain estimated from the sample covariances:

$$\tilde{\mathbf{K}} = \tilde{\mathbf{P}}^f \mathbf{H}^T (\mathbf{H} \tilde{\mathbf{P}}^f \mathbf{H}^T + \mathbf{R})^{-1} \quad (2)$$

In this expression, the superscript T denotes a matrix transpose, $\tilde{\mathbf{P}}^f$ is the approximation of the forecast error covariance matrix based on ensemble covariances and $\mathbf{R}$ is the observation error covariance matrix. A perturbed-observation analysis scheme (Burgers et al., 1998) is used to update the ensemble anomalies. No covariance localisation is applied in this work.

A 96-member ensemble is generated including perturbations of the initial conditions, winds and lateral boundary conditions. The ensemble is initialised on 13 August 00:00 using model states randomly selected among the ROMS reference solutions from 8 to 18 August at 00:00. The wind perturbation strategy aims to simulate both temporal offsets and magnitude errors potentially affecting the representation of atmospheric processes. These perturbations are generated by perturbing the bivariate Empirical Orthogonal Functions (EOFs) time series computed from the COSMO-ME model zonal and meridional wind fields. Only the 20 predominant EOFs are considered. More specifically, these perturbations consist of (i) a random amplitude increase or reduction with decorrelation time scales of 1 d and standard deviation of 50% of the reference value and (ii) a random time offset with a 6-hour standard deviation, except for the EOF associated with the daily cycle for which no temporal offset is applied. This perturbation strategy was validated at ODAS mooring where the generated perturbation variance was shown to be consistent with the actual error of the atmospheric model reference field. The boundary conditions are perturbed in a similar way, except that the bivariate EOFs are computed from the temperature and normal velocity time series provided by the MFS model. In that case, the random amplitude modification is applied with a decorrelation time scale of 3 d and the temporal offset has a 24-hour standard deviation.

2.2.2. 3-D super-ensemble. The 3DSE (Lenartz et al., 2010) provides a non-intrusive post-processing method to generate a consensual ocean forecast when multiple numerical predictions and observations are available. Individual forecasts are combined through an optimal and spatially variable weighting after interpolation on a common spatial grid. A priori spatial weight error correlations are used to update the model weights in areas lacking observations. The 3DSE analysis can be expressed using the Kalman filter formalism when defining a state vector $\mathbf{w}$ containing the spatially variable weights w of the individual models:

$$\mathbf{w} = (w_1^1 \ \cdots \ w_1^M \ w_2^1 \ \cdots \ w_N^M)^T \quad (3)$$

where N is the number of models and M is the number of ocean points in the 3DSE grid. Considering a learning period with available observations from t_{i-1} to t_i , the update step can be expressed as follows:

$$\mathbf{w}^a = \mathbf{w}^f + \hat{\mathbf{K}}(\mathbf{y} - \hat{\mathbf{H}}\mathbf{w}^f) \quad (4)$$

with the Kalman gain $\hat{\mathbf{K}}$:

$$\hat{\mathbf{K}} = \hat{\mathbf{P}}^f \hat{\mathbf{H}}^T (\hat{\mathbf{H}} \hat{\mathbf{P}}^f \hat{\mathbf{H}}^T + \mathbf{R})^{-1} \quad (5)$$

In these equations, the observation operator $\hat{\mathbf{H}}$, which projects the vectors from the space of model weights onto the observation space, contains the values of the observed variables predicted by the individual models at the times of the observations, weighted by the spatial interpolation coefficients from the 3DSE grid to the position of these observations. $\hat{\mathbf{P}}^f$ is the error covariance matrix of the model weights, including both spatial and cross-model weight error covariances. Note that the 3DSE does not enforce the model weights to sum up to one.

The present implementation of the 3DSE uses a 24-hour learning period coincident with the previously described EnKF assimilation window. The optimal weights resulting from this analysis are used to combine the individual predictions during the next forecast cycle from t_i to $t_i + 72$ hours. The method is applied recursively, i.e. $\mathbf{w}^a$ and $\hat{\mathbf{P}}^a$ are used as *a priori* weight vector and associated error covariance matrix for the successive 3DSE analysis. As for the EnKF, daily 72-hour retrospective forecasts are produced during the 11-d period of the REP10 experiment.

Although the 3DSE is expressed using a similar formalism as the EnKF and is implemented to pursue the same objective (improving ocean predictions by integrating data from models and observations), it is important to underline the fundamentally different nature of the two methods. While the EnKF incorporates the observations into a dynamical model by considering the physical model parameters as stochastic, the random variables in the 3DSE approach are the weights used to combine a limited number

of deterministic model predictions. Moreover, while the 3DSE weights persist in time during the forecast period, the EnKF state variables are propagated according to the dynamical ocean model. Therefore, no formal equivalence might be sought between the two approaches.

The 3DSE is fed by the same three high-resolution ocean models used in Mourre et al. (2012), namely the French coastal operational forecasting system PREVIMER and two regional implementations of the Navy Coastal Ocean Model (NCOM) and the ROMS, respectively. The horizontal resolution of the three models ranges between 1.5 and 2 km. Note that the ROMS model used as input for the 3DSE slightly differs from the version used in the EnKF, the major difference being the inclusion of climatological river runoffs in the latter. The interested reader can find a more detailed description of these models in Mourre et al. (2012). The 3DSE spatial domain extends from 8.5°E to 10.6°E in longitude, from 43.1°N to 44.4°N in latitude and from the surface to 200 m depth. The 3DSE grid has a 3-km horizontal resolution. The vertical discretisation is inspired from the ROMS vertical grid in the central part of the domain, with a resolution decreasing with depth, from less than 2 m at the surface to around 40 m at 200 m depth.

Distance-based Gaussian correlations are used as initial spatial weight error correlations, with horizontal and vertical decorrelation distances of 50 km and 20 m, respectively. Cross-model weight error correlations are initially set to zero. Weight error variances are initialised with a homogeneous value of 0.01 over the modelling domain. Due to the univariate formulation of the method, temperature and salinity weights are calculated independently from two separate simulations.

2.2.3. Observations. The assimilated measurement dataset is composed of: (1) surface temperature data from the OISST high-resolution daily analysis produced by the Gruppo di Oceanografia da Satellite (CNR Roma, Buongiorno Nardelli et al., 2013) and (2) subsurface temperature and salinity observations from the fleet of underwater gliders, whose trajectories are represented in Fig. 1.

The High-Resolution (6-km pixels) OISST product is obtained by spatio-temporal statistical interpolation of all available night-time measurements from available satellite infrared sensors. OISST data are incorporated only every 3 d in both systems due to the observed temporal correlations of the daily products. The OISST represents the so-called foundation surface temperature, which is assumed to correspond to the model surface temperature at 07:00, before any diurnal warming effect. Both the 3DSE and EnKF properly account for the pixel nature of OISST data by building observation operators which compute the

model averages over the OISST pixels. The OISST error variance is the sum of the analysis error provided with the OISST product, which depends on satellite passes and cloud coverage, and a fixed representation error assumed to be $(0.5^\circ\text{C})^2$.

Under the surface, every single glider profile is considered, taking into account its inclination from the vertical. These profiles are first interpolated on the vertical model grid before being incorporated in both approaches. For each level, the observed variance in the vertical grid cell is used as an approximation of the vertical representation error. The horizontal representation error variance is assumed to be $(0.25^\circ\text{C})^2$ and 0.05^2 for temperature and salinity (in practical salinity scale) measurements, respectively. The observation error covariance matrix is specified as diagonal in both approaches (i.e. $\mathbf{R}$ is exactly prescribed in the EnKF and not estimated from the ensemble of observations). To compensate for the lack of consideration of spatial error correlations, observation error variances are individually inflated by a factor equal to the number of neighbouring profiles within a model grid cell radius.

3. Comparison of the performance of the 3DSE and EnKF

The ocean temperature is first considered in this comparison. The differences are analysed based on several types of observations available for validation: (i) a central *Scanfish* section in the area of the assimilated glider measurements, (ii) a surface CTD transect covering a more extended part of the modelling domain and (iii) several CTD stations distributed over the study area. Secondly, insights into the ocean salinity forecast performance are presented.

Different statistical measures can be used to characterise the mismatch between model and observations (see, for instance Murphy, 1995 for a more detailed discussion). First, the total root-mean-square difference (RMSD) represents the average magnitude of the model-observation differences:

$$\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (m_i - o_i)^2} \quad (6)$$

where m_i and o_i are the values of the i th modelled and observed variables, respectively, and N the number of observations available for validation. The RMSD can be decomposed into two terms allowing to separate the contributions from (i) the mean difference between model and observations (hereafter referred to as *bias*) and (ii) the overall amplitude and phase disagreement between the

variable part of the two fields (hereafter referred to as *unbiased RMSD* or RMSD').

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n (m_i - o_i) \quad (7)$$

$$\text{RMSD}' = \sqrt{\frac{1}{n} \sum_{i=1}^n ((m_i - \bar{m}) - (o_i - \bar{o}))^2} \quad (8)$$

The overbar denotes the mean value. Notice the quadratic relationship between the RMSD, the bias and the unbiased RMSD:

$$\text{RMSD}^2 = \text{bias}^2 + \text{RMSD}'^2 \quad (9)$$

The unbiased RMSD can in turn be decomposed into two terms:

$$\text{RMSD}'^2 = (\sigma_m - \sigma_o)^2 + 2\sigma_m\sigma_o(1 - \text{CC}) \quad (10)$$

where σ_m and σ_o are the standard deviations of the model and the observations, respectively, and CC the linear correlation coefficient between the two fields. While the quantity $(\sigma_m - \sigma_o)$, hereafter referred to as *standard deviation difference* (SDD), compares the variability of the model to that of the observations, the correlation coefficient CC

provides a measure of the general correspondence in phase between the two fields.

3.1. Ocean temperature forecast

3.1.1. Along a central Scanfish section. A towed Scanfish vehicle equipped with CTD was deployed behind the NRV Alliance from 26 August 23:00 to 27 August 05:00. The vehicle undulated between 6 and 70 m depth. It provided high spatial resolution measurements of temperature and salinity in the central part of the study area. Figure 2 displays the Scanfish trajectory and compares the collected temperature observations to the corresponding 3DSE and EnKF predictions produced on 26 August 00:00.

Both solutions, though slightly different, provide overall agreement with the observations, with accurate temperature ranges and vertical gradients. The methods do not reproduce the measured small-scale variability (over a few kilometres), as this is beyond the capability of the numerical ocean models used in this study. The RMSD between model and observations over this section is comparable for both methods: 0.81 and 0.85°C for the EnKF and 3DSE, respectively. As a point of comparison, the control solution,

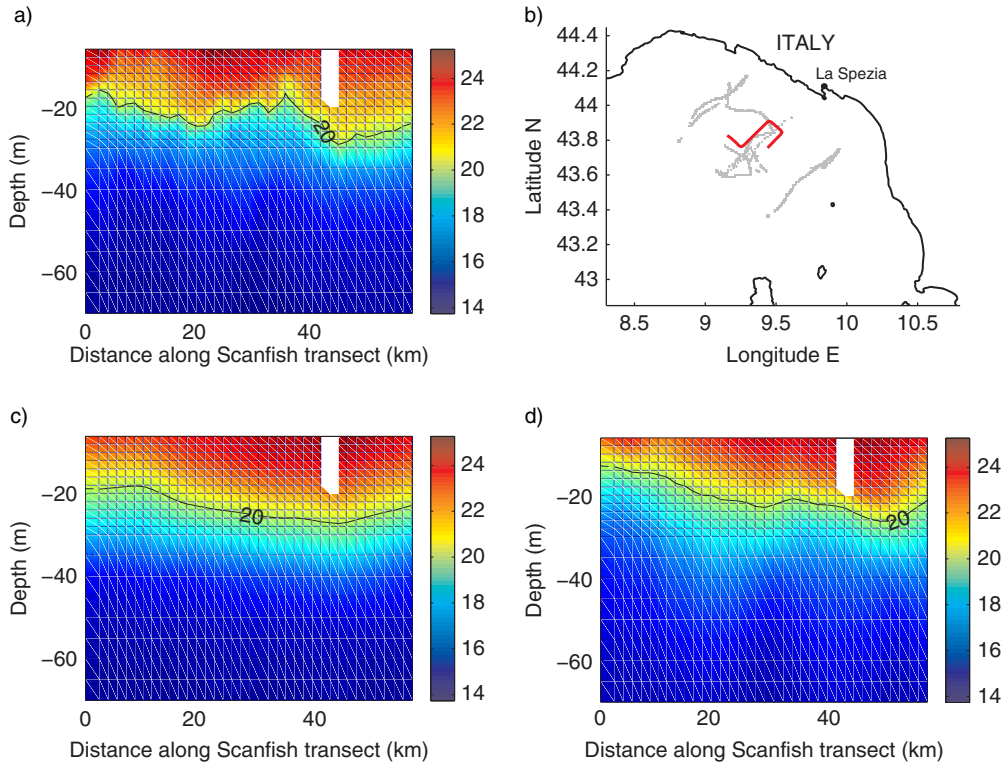


Fig. 2. Temperature section (°C) along the NRV Alliance ship route from 26 August 23:00 to 27 August 05:00: a) observed by the *Scanfish*; predicted on 26 August 00:00 by c) the 3DSE; d) the EnKF. b) In red: ship trajectory for this period (the transect starts from its westernmost position), and in grey: glider observations assimilated prior to 26 August.

which is provided here by the ensemble of ROMS simulations including the same perturbations as the EnKF but without any data assimilation, leads to a RMSD of 1.10°C .

3.1.2. At the surface. The surface temperature CTD measurements continuously collected along the Alliance route from 21 August to 3 September provide surface information over an area extending several tens of kilometres away from the position of the assimilated observations. For each daily simulation, the surface CTD data collected during the next 72-hour period are compared to their model counterparts. Figure 3 illustrates the model-data differences over the study area and the corresponding histograms. A total of 4390 model-measurement differences are computed for each predictive system over the 11 simulations from 21 to 31 August. The statistical measures associated with these observed differences are presented in Table 1.

While the absolute value of the bias is smaller for the 3DSE, the EnKF leads to a lower RMSD due to general

better pattern agreement with the observations characterised by a larger correlation coefficient. Also, the 3DSE solution is characterised by the occurrence of unlikely values evidenced by the large range (difference between the maximum and minimum values) of the differences which reaches 7.9°C (this range is 4.0°C for the EnKF solution). The heavy-tailed shape of the 3DSE histogram (Fig. 3c) illustrates these many extreme positive and negative residuals exceeding 2° in absolute value. The map of model-data differences (Fig. 3a) shows that these large 3DSE errors occur outside the area spanned by the glider measurements used to constrain the prediction. The surface temperature forecast performance of the EnKF is found to be more homogeneous over the validation area (Fig. 3b).

3.1.3. At CTD stations. A total of 65 CTD profiles were collected in the modelling domain during the REP10 experiment from 21 August to 3 September. The 3DSE and EnKF solutions are compared against each of these

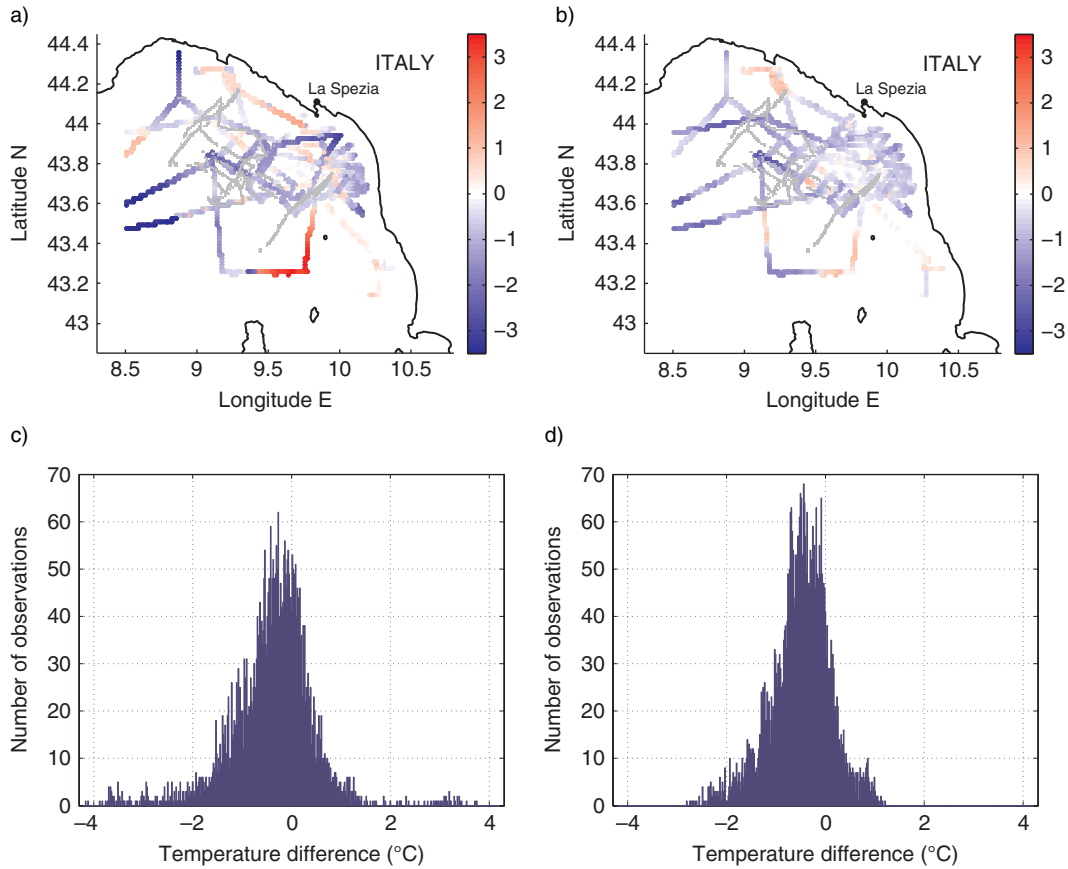


Fig. 3. Top panels: maps of temperature differences ($^{\circ}\text{C}$) between models and the ship surface CTD observations from 21 August to 3 September: a) 3DSE; b) EnKF. For each sampled location, the maximum value of the differences over the three available simulation dates is shown. In light grey: position of the assimilated observations over the whole experiment. Bottom panels: corresponding histograms for c) 3DSE and d) EnKF.

Table 1. Summary table of the statistical measures of model-observation differences

Metric	Simulation	Validation measurements		
		Ship surface CTD temperature	Subsurface CTDs	
			Temperature	Salinity
RMSD	3DSE	0.93°C	1.83°C	0.84
	EnKF	0.79°C	1.36°C	0.14
bias	3DSE	-0.38°C	-0.45°C	-0.04
	EnKF	-0.51°C	-0.43°C	0.08
RMSD'	3DSE	0.84°C	1.77°C	0.84
	EnKF	0.60°C	1.28°C	0.12
SDD	3DSE	0.26°C	-0.30°C	0.68
	EnKF	0.26°C	-0.44 °C	-0.03
CC	3DSE	0.52	0.92	0.10
	EnKF	0.78	0.96	0.68

profiles. Figure 4 shows the RMSD at each CTD station for both predictive systems. Table 2 summarises the average¹, maximum and minimum values of the RMSD over all the CTD sites. The prediction error in the neighbourhood of the assimilated glider tracks is smaller than 1.5°C for both prediction systems. Outside this area, RMSD values larger than 2°C are frequently found (in 12 of the 65 CTDs) in the 3DSE solution, especially when approaching the coast or in the non-observed southern part of the modelling domain. The maximum 3DSE RMSD over the CTD sites is larger than 5.5°C, which confirms the occurrence of out-of-range values in the 3DSE solution when moving away from the area spanned by the measurements used to constrain the method. As a consequence, the average RMSD over all CTD sites is larger for the 3DSE (1.78°C) than for the EnKF (1.25°C).

3.1.4. Summary statistics. The summary diagram (Jolliff et al., 2009) in Fig. 5 shows the absolute value for the bias, the unbiased RMSD, total RMSD and correlation coefficient obtained for both predictions over the whole period of the REP10 experiment. Surface observations from the continuous surface CTD on-board the NRV Alliance are separated from the volume measurements collected at the CTD stations. The scores of the probabilistic ROMS control run without assimilation are also displayed for comparison.

At the surface, while the 3DSE has a lower bias than the EnKF, the latter exhibits a smaller unbiased RMSD and a larger correlation than the 3DSE, as already commented in Section 3.1.2. Note that the EnKF surface bias and

unbiased RMSD are reduced compared to the control solution without assimilation. Below the surface, when considering all available model-measurement differences at CTD stations, the EnKF is found to outperform both the control ROMS simulation and the 3DSE as depicted for the four statistical measures illustrated in this diagram and detailed in Table 1. Note that the values of the correlation coefficient close to unity are indicative that the vertical temperature gradients are properly described in both model solutions.

The variability with depth of the different contributions to the total RMSD is illustrated in Fig. 6. The surface values are given by the mismatches with the surface CTD data, whereas the subsurface statistics at a given depth are computed using the model-observation differences for the entire set of CTD profiles at that depth. The largest values of the RMSD are found in the thermocline for both solutions. The performance of the EnKF is found to be superior to that of the 3DSE for all depths from the surface down to 150 m, with a marked difference in the upper 40 m layer. This difference is due to a better pattern agreement with the observations of the EnKF solution resulting from both an overestimation of the variability by the 3DSE in the upper layer, and a larger correlation coefficient of the EnKF. Note that the EnKF correlation coefficient is larger than that of the 3DSE for all sampled depths.

3.1.5. Skill scores. The following skill score (hereafter referred to as SS, Murphy, 1995) is now defined to compare the performance of both predictive models to the MEDAR/MedAtlas August climatology (MEDAR Group, 2002).

$$SS = 1 - \frac{MSD_{\text{model}}}{MSD_{\text{climatology}}} \quad (11)$$

where MSD stands for the mean-square-difference with the observations. The skill score would take the value of one in the case of a perfect match between the model and the observations. In practice, the presence of measurement errors prevents the SS to take this value even for a perfect model. A skill score ranging between one and zero indicates that the model better reproduces the observations than the climatology. Conversely, a negative SS indicates that the climatology better fits the observations than the model. Figure 7 illustrates the skill scores of the 3DSE, EnKF and control simulation for the different observation sources available for validation.

The skill score at predicting the surface temperature is larger than 0.5 for both 3DSE and EnKF, indicating a positive impact of the regular incorporation of OISST data in both systems. The EnKF has a slightly better surface SS than 3DSE. The SS at ODAS mooring (from the surface to 36 m depth), which is located in the area of the

¹Computed as the square root of the average mean square difference.

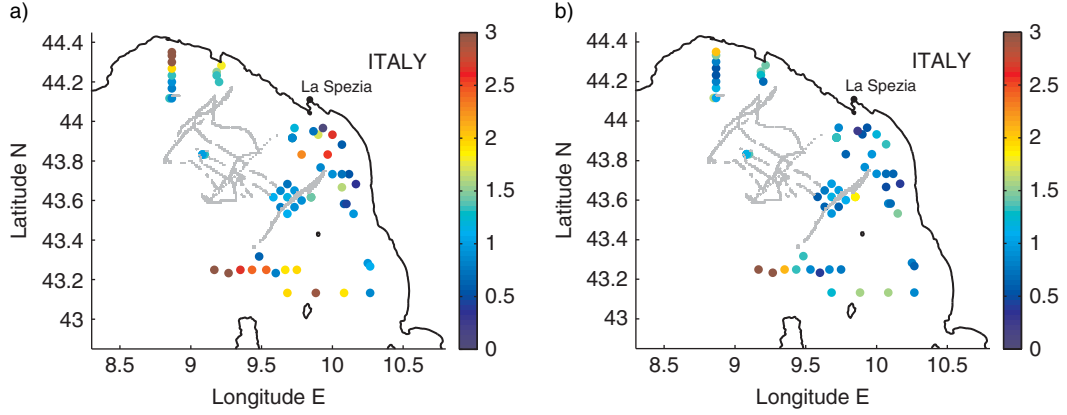


Fig. 4. Temperature RMSD ($^{\circ}\text{C}$) at CTD sites: a) 3DSE; b) EnKF. For each sampled location, the average RMSD over the three available simulation dates is shown.

assimilated glider measurements, is still very positive for both approaches, with a better performance of the 3DSE. The 3DSE also exhibits a good skill score when comparing to the glider observations collected in the first 24 hours of the forecasts. Note that these glider observations are collected by the same five platforms which collected the measurements used to constrain the solutions in the previous 24 hours. Given that the gliders travel a distance around 20 km per day, the positions of these validation measurements are necessarily very close to the observations previously incorporated in both predictive systems. While the EnKF skill is maintained approximatively constant, the 3DSE skill score is found to be significantly reduced when considering the glider data collected between 24 and 48 hours after the beginning of the forecasts. This tendency is confirmed when isolating the glider data collected in the last 24 hours of the 3-d forecast periods. In that case, the EnKF skill is found to be superior to that of the 3DSE.

Finally, the skill score based on the observations from the 65 CTD stations provides an assessment of the forecast performance over a more extended area. While the average SS of the control simulation is slightly negative as a consequence of large forecast errors for some of the CTD stations located over the continental shelf, the EnKF score is positive and comparable to that obtained when considering the glider observations. Conversely, the 3DSE score is found to be very negative, penalised by the

occurrence of very erroneous profiles in the more coastal areas and in the southern part of the modelling domain (as illustrated in Fig. 4a). Contrasting with the homogeneity of the EnKF forecast skill over the modelling domain, the 3DSE performance is found to be strongly dependent on the distance to the subsurface measurements used to constrain the method, a significant degradation being observed a few tens of kilometres away from these observations.

3.2. Ocean salinity forecast

Unlike temperature and in spite of encouraging on-going international efforts towards this objective, satellites do not yet provide regular, accurate and high-resolution observations of the ocean surface salinity variability at the regional scale. This present gap makes the prediction of salinity particularly challenging since the observations are limited to sparse in-situ profiles. From an oceanographic point of view, the upper Ligurian Sea salinity is characterised in summer by a marked subsurface minimum centred in the thermocline. This salinity minimum is associated with the presence of Modified Atlantic Water lying between the salty surface waters and the high salinity intermediate waters originating from the Levantine basin. Close to the continent, the freshwater outflows from the Arno, Serchio and Magra rivers give rise to salinity fronts extending along the coast.

Figure 8 illustrates the salinity section along the *Scanfish* transect in the central part of the modelling domain on 26 August. The observations depict this salinity minimum around 35 m depth and the salty waters at the surface. Some small-scale variability linked to intrusions of fresher water lenses at the sea surface is also observed. While the 3DSE and EnKF temperature predictions were shown to be quite similar along this section (Fig. 2c and d), the

Table 2. Properties of the temperature RMSD values over all the CTD stations

Simulation	Average RMSD ($^{\circ}\text{C}$)	Minimum RMSD ($^{\circ}\text{C}$)	Maximum RMSD ($^{\circ}\text{C}$)	No. of stations with RMSD > 2°C
3DSE	1.78	0.07	5.58	12
EnKF	1.25	0.25	3.61	4

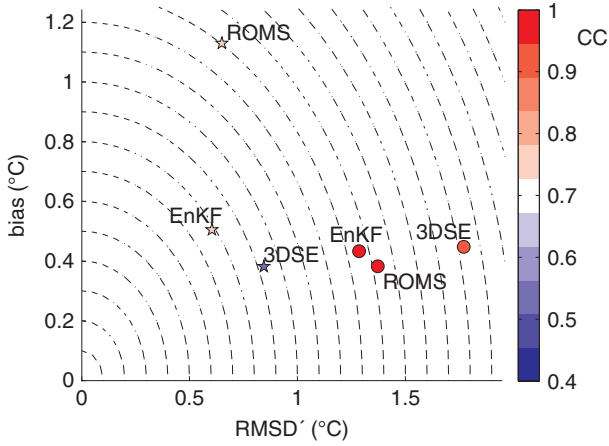


Fig. 5. Summary diagram illustrating the absolute value of the bias, the unbiased RMSD and the correlation coefficient associated with the temperature model-observation differences for the 3DSE, EnKF and control ROMS simulation. The dashed lines are isocontours of the total RMSD. The stars represent the surface model performance evaluated using the NRV Alliance surface CTD data, while the circles illustrate the subsurface performance computed from the entire set of CTD profile observations.

associated salinity forecasts exhibit more significant differences. The 3DSE reproduces the salinity minimum at a correct approximate depth. However, both the Modified Atlantic Water and the surface water are too fresh compared to the observations. The EnKF prediction also reproduces a subsurface salinity minimum, yet with other limitations. As depicted in the observations, the depth of this salinity minimum varies along the transect. However it is generally too shallow and too salty compared to the measurements. The EnKF solution is characterised by a strong positive salinity bias below the surface layer.

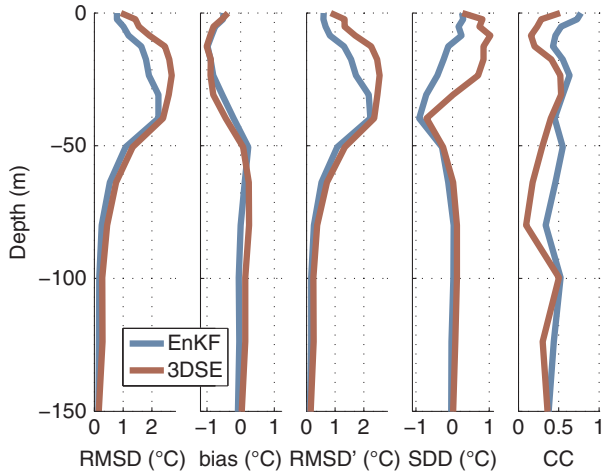


Fig. 6. Vertical variability of the statistical measures of the temperature model-observation differences over all CTD profiles.

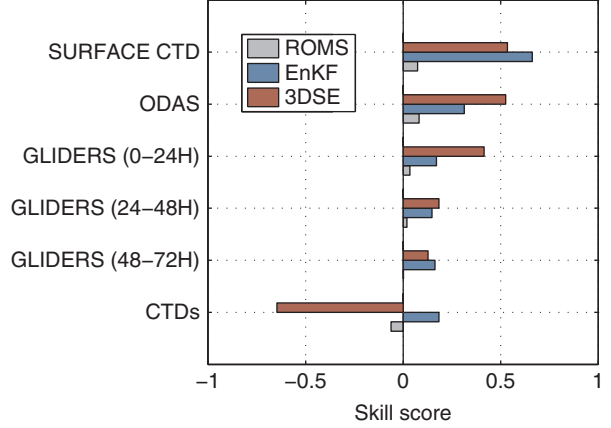


Fig. 7. Skill scores of the 3DSE, EnKF and control simulation for different validation observation sources.

Figure 9 illustrates the salinity RMSD at each of the 65 CTD stations. The error is quite homogeneous for the EnKF solution, with values smaller than 0.2 except in the shelf area south of La Spezia. This area, marked by the presence of a coastal front associated with the local river discharges, was shown to be difficult to realistically model due to the impact of small-scale phenomena (Schroeder et al., 2012). On the contrary, the 3DSE error is found to be very large for many of the profiles, especially when approaching the coast and in the southern part of the domain, as already observed with the temperature prediction. While the RMSD exceeds 0.2 at only six of the stations for the EnKF, this is the case for 46 of the 65 available CTDs when considering the 3DSE solution.

The statistical measures of the model performance for the entire set of CTD salinity measurements are summarised in Table 1. A low correlation coefficient combined with an overestimation of the variability lead to a very large unbiased RMSD for the 3DSE salinity solution. At the same time, the EnKF predictions are characterised by a much lower total RMSD with comparable contributions of the bias and unbiased RMSD. Note that the EnKF performance on salinity is only slightly improved compared to the ROMS control prediction, which has a total RMSD of 0.15 and a bias of 0.09.

4. Discussion

As described in Section 2.2.2, the 3DSE optimises the individual model weights so that the resulting model linear combination comes as close as possible to the observations under the constraint of the spatial smoothness of these weights imposed by the isotropic error correlations. Looking into greater details, the matrix $\hat{\mathbf{H}}_i \hat{\mathbf{P}}_i^f \hat{\mathbf{H}}_i^T$ in eq. (5), which represents the 3DSE model error covariances at observation points, involves the product of the weight error

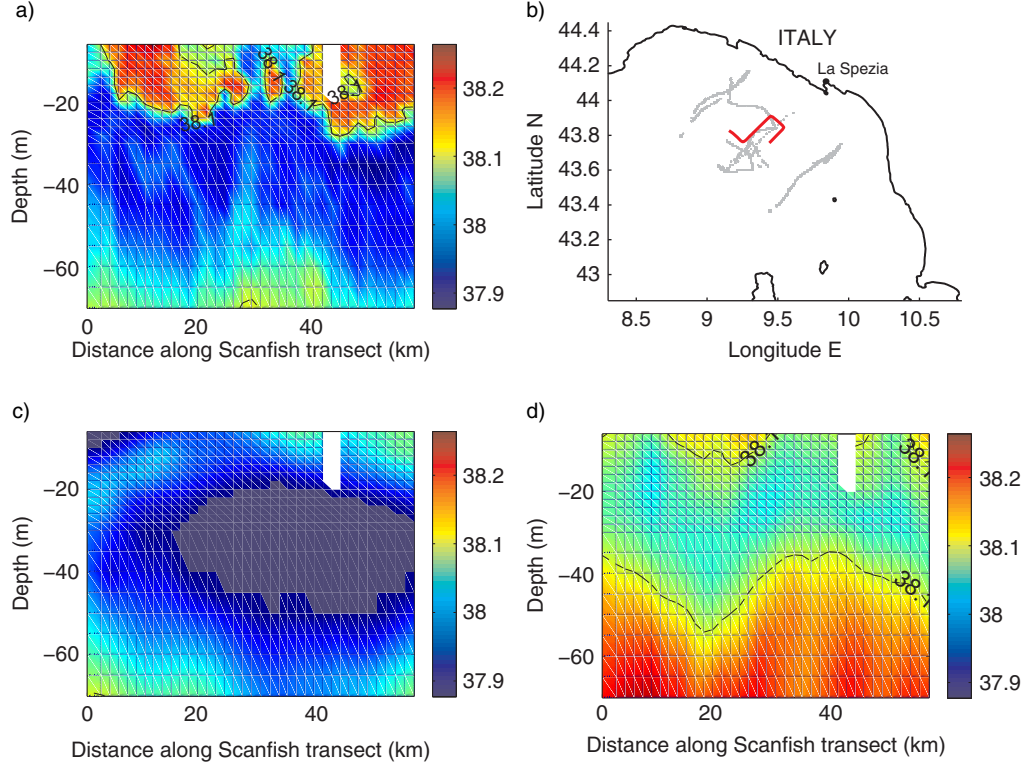


Fig. 8. Salinity section along the Scanfish transect: a) observations; c) 3DSE; d) EnKF.

covariances described in $\hat{\mathbf{P}}_i^f$ by the values of the physical variables present in $\hat{\mathbf{H}}_i$. In the absence of cross-model error correlations (as initially specified), the magnitude of the diagonal terms in $\hat{\mathbf{H}}_i \hat{\mathbf{P}}_i^f \hat{\mathbf{H}}_i^T$ (of the order of $(2^\circ\text{C})^2$ for a 20°C temperature and a 0.1 weight error standard deviation) is typically two orders of magnitude larger than the magnitude of the observation error variance [around $(0.25^\circ\text{C})^2$] during the 3DSE temperature analysis. As a consequence, the 3DSE solution is very strongly steered towards the values of the observations. Our experience is that the 3DSE

solution is generally found to almost perfectly match the constraining observations during the learning period. In this study, the near zero bias characterising the 3DSE forecast under 40 m depth in the temperature *Scanfish* transect provides an illustration of this capacity to fit the observations in an ocean layer with small variability. This explains the good performance of the 3DSE solution in the neighbourhood of the incorporated glider observations, namely in this study along the *Scanfish* transect, at the ODAS mooring and along the glider trajectories

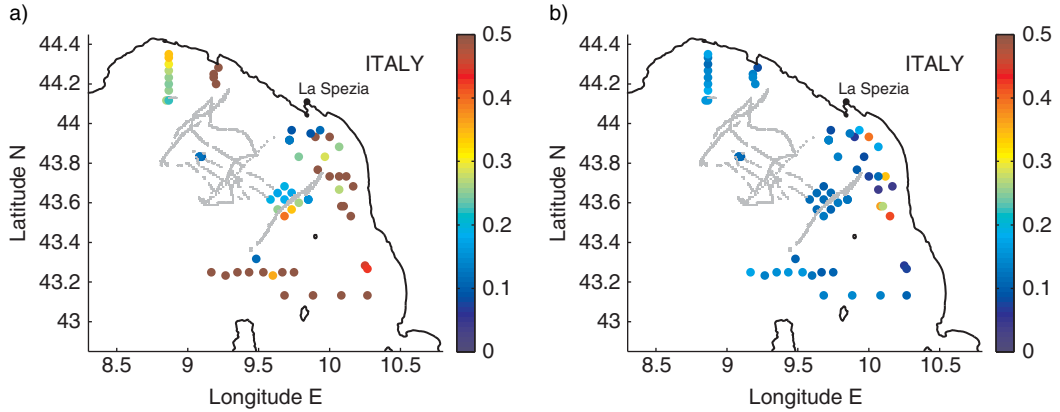


Fig. 9. Salinity RMSD at CTD sites: a) 3DSE; b) EnKF.

during the first hours of the forecast periods. In the EnKF, the error variances of the model background and observations are generally more comparable in magnitude, which results in a relative stronger constraint being exerted by the underlying numerical model and a somewhat lower capacity to fit the assimilated measurements compared to the 3DSE.

The 3DSE ocean model weights which are optimal in the observed central part of the domain can be expected to lose validity outside this area, in particular when approaching the coast where small-scale shelf processes have a large influence on the ocean fields, or in the southern part of the domain, where the local dynamics are influenced by the oceanic flows around Corsica. The high sensitivity of the 3DSE solution to small errors in the model weights (a small weight error of 0.05 being converted into a 1°C 3DSE temperature error after the product by a 20°C model value) then explains the observed frequent occurrence of very erroneous values in the 3DSE solution over these areas.

Moreover, the inferior pattern-related skills of the 3DSE reported in this study probably reflect the limited capacity of the method to create physically consistent fields. Indeed, the product of spatially variable model weights by physically-balanced model outputs breaks the dynamical balances, which remain unconstrained in the 3DSE solutions. On the contrary, the EnKF estimate, as an unweighted average of numerical ocean model solutions, better respects the underlying equations, which is reflected by larger correlation coefficients when compared to observations.

We have been investigating ways to introduce in the 3DSE physical covariances issued from the solutions of the individual numerical models, but without success. One of the difficulties comes from the fact that the whole 3DSE linear combination needs to be spatially propagated in a consistent manner, thus requiring a common weight error spatial covariance function for all individual models. Also, we haven't been able to significantly improve the 3DSE predictions by investigating different 3DSE formulations using model anomalies with respect to averaged fields instead of absolute values so as to reduce the sensitivity of the solution to errors in the model weights. An alternative formulation of the 3DSE constraining the model weights at each grid point to sum up to unity should contribute to reduce the sensitivity of the 3DSE solution to small errors in the model weights and limit the occurrence of very erroneous predictions over the modelling domain. In that case, the error would be locally bounded by the maximum error of the individual predictions. However, this constraint would also reduce the 3DSE ability to fit the constraining measurements, which is expected to deteriorate the performance of the method close to these observations.

Our findings are compatible with the 3DSE skills reported in Lenartz et al. (2010) and Mourre et al. (2012). In both studies, the 3DSE was shown to provide accurate temperature predictions using validation data including observations from the same gliders whose measurements had been considered in the previous learning period. Due to the low speed of the gliders in the water and the large amount of measurements collected by these platforms, our interpretation is that a predominant weight was given to the observations collected in the neighbourhood of the constraining measurements when computing the overall statistical performance, thus explaining the positive evaluation of the method.

The results have been extended to the prediction of salinity in the present study. The univariate formulation of the 3DSE combined with the limited amount of salinity observations and the sensitivity of the solution to small errors in the model weights, explain the occurrences of out-of-range values in the 3DSE salinity predictions outside the area spanned by the measurements used to constrain the model weights.

5. Summary

The performance of two model-data-integrative approaches of very different nature but of potential use for the operational production of short-term regional ocean predictions has been compared in this study. The multi-model 3DSE method, which benefits from the increasing number of operational ocean models available over a given region, was first considered. The 3DSE approach aims to represent the spatial variability of the individual model skills by determining spatially variable optimal model weights using distance-dependent spatial covariances to extrapolate the optimal values outside the area spanned by the incorporated measurements. The second technique considered in this study was the more conventional EnKF method implemented with a single ocean circulation model. The EnKF approach uses ensemble approximations of the multivariate and flow-dependent error covariances of the physical variables to propagate the information provided by the observations within the modelling domain. Note that the EnKF, which requires an ensemble of many tens of simulations of the numerical ocean model, demands more computing time than the 3DSE, which benefits from the availability of external model outputs, and only requires their interpolation onto a common spatial grid before performing the analysis. In our experiment, while the 3DSE daily simulation was computed in around 30 minutes using a single workstation processor, the daily 96-member EnKF computation required about 5 hours using four computing nodes, each of them providing eight

processors. However, despite the different computational costs of the two methods, both of them are affordable for operational implementations with reasonable computing resources.

The forecast skills of the 3DSE and EnKF, constrained by a common dataset composed of oceanographic observations from a fleet of five gliders and sea surface temperature information, have been compared in the Ligurian Sea. Validation measurements including temperature and salinity observations collected over the modelling domain during the REP10 experiment have allowed an evaluation of the performance of both methods.

Our results indicate that the 3DSE and EnKF perform comparably well in terms of temperature RMSD in the area spanned by the assimilated subsurface measurements. The EnKF performance was found to be quite homogeneous over the whole modelling domain and consistently better than both the climatology and the control solution without assimilation. On the contrary, the performance of the 3DSE was shown to deteriorate outside the area spanned by the constraining measurements, with the occurrence of unrealistic values in the more coastal zones or in areas under the influence of specific dynamical processes. While the absolute temperature bias had a similar magnitude in both simulations, the overall pattern agreement with the observations, as represented by the unbiased RMSD, was significantly better in the EnKF solution, especially in the upper 40 m where the variability is enhanced. Considering salinity, whereas the EnKF and 3DSE were providing a similar performance in the neighbourhood of the assimilated measurements, the 3DSE produced large salinity errors outside this area, which did not occur with the EnKF except near the eastern Ligurian coastal front.

The high sensitivity of the 3DSE solution to small errors in the individual model weights combined with the lack of physical background for the spatial propagation of these weights into non-observed areas, which leads to the occurrence of large 3DSE errors, constitutes a significant limitation of this multi-model approach. Despite the relatively low cost of the method and its good prediction skill in the neighbourhood of the measurements used to constrain the solution, the deteriorated performance outside this area limits its use for practical oceanographic applications with scarce subsurface observations. The flow-dependent spatial and multivariate error covariances used in the EnKF allow a more realistic propagation of the information from the position of the assimilated observations to the whole modelling domain, especially in regional seas characterised by small-scale, local and anisotropic dynamical processes.

6. Acknowledgements

We thank the Italian Air Force National Meteorological Center (Centro Nazionale per la Meteorologia e Climatologia Aeronautica, CNMCA) for providing COSMO-ME data, EU FP7 EUROSITES for providing the observations at ODAS ITALIA 1 mooring, Alexander Barth (University of Liège) for sharing the ROMS Ligurian Sea model grid and the Istituto Nazionale di Geofisica e Vulcanologia Bologna for delivering the MFS predictions. We also thank the institutions providing the ocean forecasts used as input for the 3DSE: IFREMER-SHOM-Meteo France for the PREVIMER system and the Naval Research Laboratory–Stennis Space Center for the NCOM model. We are grateful to Chuck Trees for the scientific lead of the REP10 experiment, and Richard Stoner, Daniele Cecchi, Giuliana Pennucci, Gisella Baldasserini, Craig Lewis, Matt Cofin and Domencio Galletti for maintaining and piloting the gliders. The North Atlantic Treaty Organization funded this work.

References

- Alvarez, A., Chiggiato, J. and Mourre, B. 2013. Under the sea: rapid characterization of restricted marine environments. *IEEE Robot. Autom. Mag.* **20**(3), 42–49.
- Bonavita, M. and Torrisi, L. 2005. Impact of a variational objective analysis scheme on a regional area numerical model: the Italian air force weather service experience. *Meteorol. Atmos. Phys.* **88**, 39–52.
- Bonavita, M., Torrisi, L. and Marcucci, F. 2010. Ensemble data assimilation with the CNMCA regional forecasting system. *Q. J. Roy. Meteorol. Soc.* **136**(646), 132–145.
- Brasseur, P., Bahurel, P., Bertino, L., Birol, F., Brankart, J.-M. and co-authors. 2005. Data assimilation in operational ocean forecasting systems: the MERCATOR and MERSEA developments. *Q. J. Roy. Meteorol. Soc.* **131**, 3561–3582.
- Buongiorno Nardelli, B., Tronconi, C., Pisano, A. and Santoleri, R. 2013. High and ultra-high resolution processing of satellite Sea Surface Temperature data over Southern European Seas in the framework of MyOcean project. *Rem. Sens. Environ.* **129**, 116.
- Burgers, G., Van Leeuwen, P. J. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126**, 1719–1724.
- Chao, Y., Li, Z., Farrara, J., McWilliams, J. C., Bellingham, J., Capet, X. and co-authors. 2009. Development, implementation and evaluation of a data-assimilative ocean forecasting system off the central California coast. *Deep-Sea Res II*. **56**, 100–126.
- Doblas-Reyes, F., Hagedorn, R. and Palmer, T. N. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. *Tellus A*. **57**, 234–252.

- Dombrowsky, E., Bertino, L., Brassington, G. B., Chassignet, E. P., Davidson, F. and co-authors. 2009. GODAE systems in operation. *Oceanography*. **22**(3), 8095.
- Dowell, D. C., Zhang, F., Wicker, L. J., Snyder, C. and Crook, N. A. 2004. Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: ensemble Kalman filter experiments. *Mon. Weather Rev.* **132**, 19822005.
- Evensen, G. 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynam.* **53**, 343–367.
- Haidvogel, D. B., Arango, H., Budgell, W. P., Cornuelle, B. D., Curchitser, E. and co-authors. 2008. Ocean forecasting in terrain-following coordinates: formulation and skill assessment of the Regional Ocean Modeling System. *J. Comput. Phys.* **227**, 3595–3624.
- Houtekamer, P. L., Mitchell, H. L. and Deng, X. 2009. Model error representation in an operational ensemble Kalman filter. *Mon. Weather Rev.* **137**, 2126–2143.
- Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. and co-authors. 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *J. Mar. Syst.* **76**, 64–82.
- Keppenne, C. L., Rienecker, M., Kurkowski, N. P. and Adamec, D. A. 2005. Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction. *Nonlinear Processes Geophys.* **12**, 491–503.
- Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z. and co-authors. 1999. Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*. **285**(5433), 1548–1550.
- Leeuwenburgh, O. 2007. Validation of an EnKF system for OGCM initialization assimilating temperature, salinity, and surface height measurements. *Mon. Weather Rev.* **135**, 125–139.
- Lenartz, F., Mourre, B., Barth, A., Beckers, J.-M., Vandenbulcke, L. and co-authors. 2010. Enhanced ocean temperature forecast skills through 3-D super-ensemble multi-model fusion. *Geophys. Res. Lett.* **37**(19), L19606.
- Logutov, O. G. and Robinson, A. R. 2005. Multimodel fusion, error parameter estimation. *Q. J. Roy. Meteorol. Soc.* **131**, 3397–3408.
- Marchesiello, P., McWilliams, J. C. and Shchepetkin, A. 2001. Open boundary conditions for long-term integration of regional oceanic models. *Ocean Model.* **3**, 1–20.
- Marta-Almeida, M., Pereira, J. and Cirano, M. 2011. Development of a pilot Brazilian regional operational ocean forecast system, REMO-OOF. *J. Oper. Oceanogr.* **4**(2), 3–15.
- MEDAR Group. 2002. MEDATLAS 2002, Mediterranean and Black Sea database of temperature salinity and bio-chemical parameters and climatological atlas, *IFREMER edition (4 CD-Roms)*.
- Mourre, B. and Alvarez, A. 2012. Benefit assessment of glider adaptive sampling in the Ligurian Sea. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **68**, 68–78.
- Mourre, B., Chiggiato, J., Lenartz, F. and Rixen, M. 2012. Uncertainty forecast from 3-D super-ensemble multi-model combination: validation and calibration. *Ocean Dynam.* **62**(2), 283–294.
- Murphy, A. H. 1995. The coefficients of correlation and determination as measures of performance in forecast verification. *Weather Forecast.* **10**, 681–688.
- Oke, P. R., Brassington, G. B., Griffin, D. A. and Schiller, A. 2008. The Bluelink ocean data assimilation system (BODAS). *Ocean Model.* **21**, 46–70.
- Pinardi, N. and Coppini, G. 2010. Operational oceanography in the Mediterranean Sea: the second stage of development. *Ocean Sci.* **6**, 263–267.
- Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M. 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* **133**, 1155–1174.
- Rixen, M., Book, J. W., Carta, A., Grandi, V., Gualdesi, L. and co-authors. 2009. Improved ocean prediction skill and reduced uncertainty in the coastal region from multimodel superensembles. *J. Mar. Syst.* **78**, S282–S289.
- Sakov, P., Counillon, F., Bertino, L., Lisaeter, K. A., Oke, P. R. and co-authors. 2012. TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci.* **8**, 633–656.
- Sakov, P., Evensen, G. and Bertino, L. 2010. Asynchronous data assimilation with the EnKF. *Tellus*. **62**, 24–29.
- Schroeder, K., Chiggiato, J., Haza, A. C., Griffa, A., Özgökmen, T. M. and co-authors. 2012. Targeted Lagrangian sampling of submesoscale dispersion at a coastal frontal zone. *Geophys. Res. Lett.* **39**, L11608.
- Shchepetkin, A. F. and McWilliams, J. C. 2005. The Regional Oceanic Modeling System (ROMS): a split explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Model.* **9**, 347–404.
- Wang, X., Chao, Y., Thompson, D. R., Chien, S. A., Farrara, J. and co-authors. 2013. Multi-model ensemble forecasting and glider path planning in the Mid-Atlantic Bight. *Continental Shelf Res.* **63**, S223–S234.
- Zodiatis, G., Lardner, R., Hayes, D. R., Georgiou, G., Sofianos, S. and co-authors. 2008. Operational ocean forecasting in the Eastern Mediterranean: implementation and evaluation. *Ocean Sci.* **4**(1), 31–47.