SCIENCE AND TECHNOLOGY ORGANIZATION
CENTRE FOR MARITIME RESEARCH AND EXPERIMENTATION

**Reprint Series** **CMRE-PR-2019-002**

# The impact of covariance localization on the performance of an ocean EnKF system assimilating glider data in the Ligurian Sea

Silvia Falchetti, Alberto Alvarez

May 2019

Originally published in:

Journal of Marine Systems, vol 180, April 2018, pp 76-89,

doi: doi.org/10.1016/j.jmarsys.2017.12.008

# About CMRE

The Centre for Maritime Research and Experimentation (CMRE) is a world-class NATO scientific research and experimentation facility located in La Spezia, Italy.

The CMRE was established by the North Atlantic Council on 1 July 2012 as part of the NATO Science & Technology Organization. The CMRE and its predecessors have served NATO for over 50 years as the SACLANT Anti-Submarine Warfare Centre, SACLANT Undersea Research Centre, NATO Undersea Research Centre (NURC) and now as part of the Science & Technology Organization.

CMRE conducts state-of-the-art scientific research and experimentation ranging from concept development to prototype demonstration in an operational environment and has produced leaders in ocean science, modelling and simulation, acoustics and other disciplines, as well as producing critical results and understanding that have been built into the operational concepts of NATO and the nations.

CMRE conducts hands-on scientific and engineering research for the direct benefit of its NATO Customers. It operates two research vessels that enable science and technology solutions to be explored and exploited at sea. The largest of these vessels, the NRV Alliance, is a global class vessel that is acoustically extremely quiet.

CMRE is a leading example of enabling nations to work more effectively and efficiently together by prioritizing national needs, focusing on research and technology challenges, both in and out of the maritime environment, through the collective Power of its world-class scientists, engineers, and specialized laboratories in collaboration with the many partners in and out of the scientific domain.



**NOTE:** The CMRE Reprint series reprints papers and articles published by CMRE authors in the open literature as an effort to widely disseminate CMRE products. Users are encouraged to cite the original article where possible.

Contents lists available at ScienceDirect

Journal of Marine Systems

journal homepage: www.elsevier.com/locate/jmarsys

# The impact of covariance localization on the performance of an ocean EnKF system assimilating glider data in the Ligurian Sea

Silvia Falchetti[a], Alberto Alvarez[b,*]

[a] NATO Science & Technology Organization (STO), Centre for Maritime Research and Experimentation (CMRE), Viale San Bartolomeo 400, 19126 La Spezia, Italy
[b] Dept. of Marine Ecology (MARE), Instituto Mediterraneo de Estudios Avanzados-IMEDEA (CSIC-UIB), Miquel Marques 21, 07190 Esporlas, Spain

ARTICLE INFO

*Keywords:*
Data assimilation
Ensemble Kalman filter
Predictability
Regional ocean predictions
Underwater gliders
Ocean observing networks

ABSTRACT

Data assimilation through an ensemble Kalman filter (EnKF) is not exempt from deficiencies, including the generation of long-range unphysical correlations that degrade its performance. The covariance localization technique has been proposed and used in previous research to mitigate this effect. However, an evaluation of its performance is usually hindered by the sparseness and unsustained collection of independent observations.

This article assesses the performance of an ocean prediction system composed of a multivariate EnKF coupled with a regional configuration of the Regional Ocean Model System (ROMS) with a covariance localization solution and data assimilation from an ocean glider that operated over a limited region of the Ligurian Sea. Simultaneous with the operation of the forecast system, a high-quality data set was repeatedly collected with a CTD sensor, i.e., every day during the period from 5 to 20 August 2013 (approximately 4 to 5 times the synoptic time scale of the area), located on board the *NR/V Alliance* for model validation. Comparisons between the validation data set and the forecasts provide evidence that the performance of the prediction system with covariance localization is superior to that observed using only EnKF assimilation without localization or using a free run ensemble. Furthermore, it is shown that covariance localization also increases the robustness of the model to the location of the assimilated data. Our analysis reveals that improvements are detected with regard to not only preventing the occurrence of spurious correlations but also preserving the spatial coherence in the updated covariance matrix. Covariance localization has been shown to be relevant in operational frameworks where short-term forecasts (on the order of days) are required.

## 1. Introduction

An accurate estimation of the ocean state is required for many interdisciplinary applications, including acoustic, biological, physical and optical sciences and technologies (Lermusiaux et al., 2006). The dynamics of oceanic processes are nonlinear and highly variable, and they involve interactions across several temporal and spatial scales. These characteristics, which form the basis of the non-deterministic nature of ocean predictions, make such estimations unique and require sophisticated numerical ocean models (Brasseur, 2006). These systems are affected by an intrinsic predictability limit (Lorenz, 1969) with an inherent scale connected to the nonlinearity of the dynamical equations and to the errors in the initial conditions (Robinson and Sellschopp, 2002), the atmospheric forcing and the boundary conditions.

Data assimilation techniques aim to estimate the state of the ocean and the associated uncertainties as accurately as possible by integrating observations into ocean model states while using consistency constraints that obey the dynamical principles governing the observed system. Variational and sequential methods are among the most widely used data assimilation schemes for regional ocean systems, and there have been exciting recent advances in ensemble and four-dimensional variational approaches (Edwards et al., 2015). In particular, the ensemble Kalman filter (EnKF) (Evensen, 1994) has drawn increasing attention due to its ease of implementation and its ability to forecast ocean states and their corresponding uncertainties. Unlike the traditional Kalman filter, the EnKF can operate using nonlinear models by forecasting an ensemble of states to compute an ensemble mean and covariance, from which a single Kalman gain is derived. An analysis of each member of the ensemble allows for a mean analysis to be derived. Interested readers are referred to Evensen (2003) for further details on the EnKF approach.

Although EnKF data assimilation systems have been used in several real-world applications with state-of-the-art atmospheric models (Whitaker et al., 2008; Houtekamer et al., 2005) and ocean models at both global (Keppenne et al., 2005; Zhang et al., 2007) and regional scales (Mourre and Chiggiato, 2014), the implementation of such a

* Corresponding author.
*E-mail addresses:* alberto.alvarez@imedea.uib-csic.es, alberto.alvarez@imedea.uib.es (A. Alvarez).

https://doi.org/10.1016/j.jmarsys.2017.12.008
Received 18 April 2017; Received in revised form 19 December 2017; Accepted 31 December 2017
Available online 04 January 2018
0924-7963/ 

system is not free of difficulties. The limited statistical representativeness of the ensemble is probably the most fundamental issue. This issue originates from the small size of the ensemble when compared with the number of states accessible to the dynamics of the ocean. Inbreeding (Houtekamer and Mitchell, 2001) and the development of long-range spurious correlations in the ensemble covariance (Anderson, 2001) may be caused by a statistical misrepresentation of the ensemble. The former refers to an underestimation of the analysis error covariance after each assimilation cycle, while the latter refers to the unphysical correlations between distant locations generated by the forecast covariance. Covariance inflation (Anderson and Anderson, 1999) and localization (Houtekamer and Mitchell, 2001) have been proposed to mitigate the effects of inbreeding and spurious correlations, respectively.

Data assimilated into ocean prediction models are collected using ocean observing systems. In conjunction with autonomous profiling floats, glider technology is being used to transform ocean observing technologies from individual platform-based designs to networks of sensor nodes. This observational approach has been integrated into many current ocean observatories, such as the US Integrated Ocean Observing System (IOOS) and the Australian Integrated Marine Observing System (IMOS). Gliders make use of buoyancy changes and utilize their low-drag hydrodynamic shapes to perform zig-zag motions between the surface and the bottom of the ocean, inducing a net horizontal displacement. Their nominal speed is approximately 0.5 m $s^{-1}$ with spatial cycle periods that depend on the programmed pitch and immersion depths. Thanks to a buoyancy-based propulsion mechanism, the endurance of a glider can reach up to several months. The maneuverability of a glider, although limited to some degree by the strength of the velocity field, is another advantage offered by this technology.

A few studies in the literature have investigated the exploitation of gliders in operational ocean forecasting systems. A remarkable case was provided by the Autonomous Ocean Sampling Networks (AOSN)-II field experiment conducted in Monterey Bay, California, in 2003 (Ramp et al., 2009). The day-to-day physical variability in that ocean region was predicted using different ocean prediction models that assimilated data from heterogeneous observational assets, including a glider fleet (Chao et al., 2008; Lermusiaux, 2007; Shulman et al., 2009). The impact of the assimilation of glider observations improved the short-term (1–1.5 days) subsurface salinity and surface temperature forecasts (Shulman et al., 2009); meanwhile, for more extended forecasts, accurate atmospheric forcing data play a critical role. Further research has also demonstrated an improvement in the forecasting skills of ocean prediction models with the assimilation of glider observations of temperature and salinity fields (e.g., Zhang et al., 2010a, 2010b; Jones et al., 2012; Gangopadhyay et al., 2013; Mourre and Chiggiato, 2014).

The importance of assimilating all available information from gliders, including the vertically averaged velocity in addition to salinity and temperature observations, was emphasized by Dobricic et al. (2010) when they used an operational forecasting model for the Ionian Sea (eastern Mediterranean Sea). Mourre and Alvarez (2012) exploited the autonomy and maneuverability of gliders to explore the benefits of piloting a glider with an adaptive sampling procedure in the western Ligurian Sea under a fully operational framework. This adaptivity required a continuous feedback of information between the glider and an operational forecasting system based on a 3D super-ensemble (3DSE) assimilation technique (Lenartz et al., 2010). The above studies, among others, suggested that glider observations could significantly contribute to improving the performance of present-day operational forecasting systems.

An evaluation of the performance of an operational ocean forecast system is an important aspect of the system's development and exploitation. Determining the extent to which the events predicted by the model will compare to a corresponding set of independently obtained and reliable observations, is the most appropriate method to evaluate the performance of an operational forecasting system (Willmott et al., 1985). While the statistical measures of a model's performances are well defined, the difficulties in assessing those performances are often generated by the sparseness and unsustained collection of independent observations due to economic and/or operational limits generally encountered when observations are needed to support data assimilation and model validation tasks. As a consequence, a coherent view of the spatiotemporal structure of the prediction error is not usually well defined.

This article investigates the impact of the covariance localization on an EnKF used to assimilate glider data by means of a field experiment conducted by the Centre for Maritime Research and Experimentation (CMRE) in an area of the Ligurian Sea (western Mediterranean Sea). The trial's specific objectives include an assessment of the predictive capabilities of a multivariate EnKF, which is augmented by covariance localization and assimilated data from a glider, coupled with a regional configuration of the Regional Ocean Model System (ROMS) (Mourre and Chiggiato, 2014; Falchetti et al., 2015). A high-quality data set was repeatedly collected each day over a period (15 days) that is longer than the synoptic time scale in the region (3–4 days, Alvarez and Mourre, 2012) to allow for unique spatiotemporal tracking of the error forecasts.

The article is organized as follows. The methodology used in the EnKF system and the observational data set employed in this study are described in Section 2. The results obtained from the forecast system and the ocean circulation pattern observed in the region are described in Section 3. Finally, Section 4 provides a discussion and concludes the study.

## 2. Data and methods

### 2.1. The Mediterranean Rapid Environmental Picture 2013 (MED-REP13)

During 5–20 August 2013, a field experiment known as the Mediterranean Rapid Environmental Picture 2013 (MED-REP13) was conducted by the CMRE in a nearly rectangular area (90 km by 70 km in the along- and cross-shore directions, respectively) of the northeastern Ligurian Sea (western Mediterranean Sea, Fig. 1). The main scope of this field experiment was to investigate the operational feasibility and benefits of using a heterogeneous ocean observing network to characterize the marine environment.

The seafloor depths range from approximately 50 m to almost 1800 m in this area. Oceanographically, the region is characterized by a current system called the Northern Current (NC). This current flows northward at speeds of 0.3–0.4 m $s^{-1}$ along the continental slope, which is approximately 20–35 km off the shore of the Italian coast. The current extends down to a depth of 300 m. Below the mixed layer, the current is characterized by low salinities (37.8–38.3) and warm (14–16 °C) water masses corresponding to the salinity and temperature signatures of the Modified Atlantic Waters (MAW), which result from the inflow through the Strait of Gibraltar (Millot, 1999; Schroeder et al., 2008). The regional circulation pattern is subjected to significant dynamical variabilities reflected in the meandering nature of the NC and the presence of intense eddy activities (Marullo et al., 1985).

A Slocum glider (Webb et al., 2001) named Jade was operated in this region for the duration of the field experiment. From 5 to 10 August, the glider transited the northern part of the area before it turned towards the southern portion, which it traversed from 15 to 19 August (Fig. 1). The glider trajectory was the result of a compromise between scientific requirements and operational needs. In total, the glider conducted 285 dives to measure the conductivity, temperature and depth between 20 and 200 m using a pumped Seabird 41 CTD sensor operating at 0.5 Hz. The glider was programmed to surface every 3 h to transmit the collected data. Based on the climatological means and standard deviations of the temperature and salinity in the region, the outliers in the data were identified and removed; before calculating the salinity and potential density, the CTD data were low-pass filtered and corrected for cell thermal mass effects. Subsequently, the density

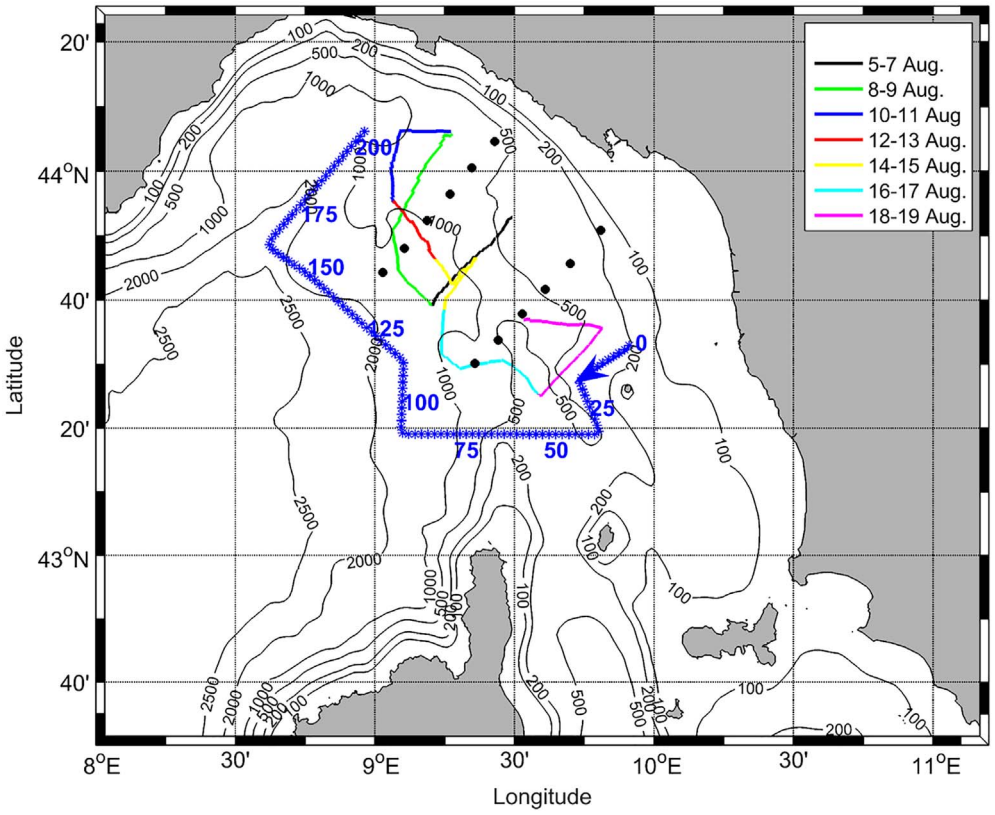


**Fig. 1.** Real trajectories for the glider Jade during 5–7 August (black line), 8–9 August (green line), 10–11 August (blue line), 12–13 August (red line), 14–15 August (yellow line), 16–17 August (cyan line) and 18–19 August (magenta line). The CTD locations are represented by black circles. The northern CTD transect was sampled on the 8, 10, 12, 16, 17, 18, and 19 of August, while the southern CTD transect was sampled on the 9, 11, 13,14, 15, 16, 17, 18 and 19 of August. The blue-star line denotes the track of the towed ScanFish MKII vehicle; the blue labels mark the horizontal track every 25 km from the starting sampling point located northward of the Gorgona island. The thin black lines denote the bathymetry contours for the 100 m, 200 m, 500 m, 1000 m, 2000 m and 2500 m depth levels as defined in the GEBCO One Minute Grid dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

inversions were identified and corrected. Finally, the time series of the glider measurements were interpolated into vertical profiles with a vertical spacing of 1 m. A total of 570 vertical profiles of the temperature and salinity with a horizontal interval of ~500 m were acquired after applying these data quality control measures.

For the purpose of collecting a validation dataset for the performance assessment, the oceanographic conditions within the study region were sampled daily from 8 to 20 August using nighttime (daylight time was devoted to acoustic operations) conductivity, temperature and depth (CTD) casts with an SBE911 from the *NRV Alliance* (Fig. 1). The CTD samples were distributed along two transects spaced approximately 40 km apart. The two parallel tracks were oriented perpendicular to the shelf break at depths ranging from 100 to 1000 m. The CTD stations were spaced approximately 10 km from each other and reached a maximum sampling depth of 600 m, with the exception of one CTD cast that sampled down to a depth of 1000 m to characterize the deepest layers. The northern CTD transect was sampled seven times on 8, 10, 12, 16, 17, 18, and 19 August, while the southern CTD transect was sampled nine times on 9, 11, 13,14, 15, 16, 17, 18 and 19 August. The different sampling frequencies were purely dictated by operational constraints.

Temperature and salinity measurements were gathered along the periphery of the surveyed domain on 13 August from a towed ScanFish MKII vehicle equipped with a SeaBird 49 CTD (Fig. 1). The instrument followed a vertical sawtooth trajectory between depths from 5 to 240 m with horizontal and vertical speeds of 3 and 1 m $s^{-1}$, respectively, allowing for a horizontal sampling resolution of approximately 3 km. These data are employed to analyze potential biases in the model runs introduced through the open boundaries. All the in situ measurements of salinity are reported using the Practical Salinity Scale (Unesco, 1985).

The CTD data were complemented with current measurements acquired along the ship track using a vessel-mounted, 75-kHz Ocean Surveyor Acoustic Doppler Current Profiler (ADCP) during the field experiment. The ADCP was configured to measure the current at 40 depth intervals ranging between 30 m and 650 m. The depth layers were equally spaced at every 16 m. The ADCP data were processed and averaged to obtain one measurement for every 2 min. Further details on the ADCP data collection and processing schemes are described in Borrione et al. (2016).

Satellite data were not considered in the present study. As was previously reported in Borrione et al. (2016), limitations in the satellite data coverage (partially due to cloud cover during the field experiment and to the low quality of the satellite imagery) as well as the poor surface thermal signatures of the subsurface water masses prevented a fruitful exploitation of sea surface temperature (SST) imagery. Moreover, no altimetry track transects passed over the computational domain, while the delayed-time gridded maps from Aviso (http://aviso.oceanobs.com/duacs/) were too coarse (1/8° in the Mediterranean Sea) to resolve the small (on the order of 10–100 km) coastal features in the region of interest (Birol et al., 2010; Borrione et al., 2016).

The performance of an EnKF data assimilation system while forecasting a temperature field was previously analyzed in a field experiment conducted in the region (Mourre and Chiggiato, 2014). The limited performance of this EnKF data assimilation system observed during the initial model runs while additionally assimilating salinity data motivated the focus on this physical variable. Moreover, the salinity signature of the MAW in the region provides a natural tracer for tracking the local circulation. The effects of local sources/sinks in the salinity (rain and river runoffs) were negligible in this area during the period of the experiment, thereby reinforcing the tracer nature of the salinity signature. For these reasons, only the salinity field will be considered in this work for validation purposes.

### *2.2. The numerical configuration of the Regional Ocean modeling system (ROMS)*

The circulation in the Ligurian Sea was simulated using a regional configuration of the ROMS. The ROMS is a primitive-equation, finite-difference, hydrostatic and free-surface ocean model that uses

generalized, terrain-following vertical s-coordinates (Haidvogel et al., 2008). The model domain covers the entire Ligurian Sea with two open boundaries on the western and southern sides located at 8°E and 42.5°N, respectively. The horizontal resolution of the model grid is approximately 1.8 km. The vertical grid uses 32 vertical levels that are non-linearly stretched to allow for a finer resolution of the surface boundary layer. The surface fluxes are calculated interactively every 3 h via bulk formulations using the atmospheric fields of the air temperature, humidity, winds and cloud cover from the 7-km-resolution Mediterranean configuration of the atmospheric model from the Consortium for Small-scale Modeling (COSMO-ME) operated by the Italian Air Force National Meteorological Center.

This model is one-way nested in the Mediterranean Forecast System (MFS; Pinardi and Coppini, 2010). The boundary conditions used here are described in Mourre and Chiggiato (2014). Radiation and nudging boundary conditions are applied at the boundaries for baroclinic flows and tracers. Barotropic flows are prescribed by means of Flather boundary conditions (Flather, 1976), while Chapman boundary conditions (Chapman, 1985) are adopted for the free surface. A generic length-scale vertical mixing parameterization scheme (Umlauf and Burchard, 2003) is activated using the Kantha and Clayson stability functions (Kantha and Clayson, 1994). A third-order up-stream bias advection scheme with velocity-dependent hyperviscosity is used to solve the nonlinear terms in the horizontal plane, while a fourth-order centered scheme is used for the vertical advection. No explicit viscosity is used, and the horizontal diffusion is modeled using a Laplacian operator along the geopotential surfaces.

### 2.3. The assimilation configuration

#### 2.3.1. The ensemble Kalman filter (EnKF)

An asynchronous multivariate EnKF formulation was used to assimilate the salinity and temperature observations at a time different from the time of the analysis (Mourre and Chiggiato, 2014). This advanced sequential data assimilation scheme uses an ensemble of the perturbed model simulations (here, consisting of 100 members) to approximate the model error covariance and the spatiotemporal variability. The algorithm advances in two steps: an analysis step and a forecast step.

Given the vector of the glider observations $y_i$ collected between $t_i$-48 h and $t_i$ and an ensemble of the model forecasts simulating this period, the present implementation of the method produces an analysis of the ensemble of model states at the time $t_i$. The augmented state vector $x_i$ contains the model state variables (i.e., the temperature, salinity, meridional and zonal velocity components and sea surface height) at the time $t_i$ and the model values at the position and time of the observations collected during the preceding 48 h. The following equation is used to update the ensemble mean state vector:

$$\overline{x}_i^a = \overline{x}_i^f + K_i(y_i - H_i\overline{x}_i^f) \tag{1}$$

where the overbars denote the ensemble mean, the superscripts $f$ and $a$ represent the forecast and analysis states, respectively, $H_i$ is the observational operator projecting the model state onto the observational space, and $K_i$ is the Kalman gain, which takes into account the observational error $R_i$ and the background error covariance $P_i^f$ to determine the extent to which observations are weighted relative to the background. The value of $K_i$ is calculated as follows:

$$K_i = P_i^f H_i^T (H_i P_i^f H_i^T + R_i)^{-1} \tag{2}$$

In this expression, the superscript $T$ denotes a matrix transpose. Given the analysis ensemble, the forecast step simply involves forecasting each member forward in time to the point at which the next observations are available (here, 48 h).

The ensemble initialization technique follows the approach described in Falchetti et al. (2015). This procedure involves only perturbations of the initial conditions to achieve the initial ensemble spread. The ensemble members are generated by taking snapshots of the forecast/analysis state values during July, August and September (at approximately every ten days) from the MFS 10-year reanalysis. The ocean state anomalies are computed by removing the ensemble mean from the ensemble members, and the final perturbed members are obtained by adding the daily MFS ocean state forecasted at the initialization date (5 August) to each anomaly state. To ensure dynamic stability and correct multivariate correlations (Evensen, 2003), a 48-hour spin-up time is allowed before commencing the assimilation.

In general, without a way to introduce realistic forecasting errors into the ensemble projection, the ensemble spread will tend to collapse (Turner et al., 2008). One way to account for unrepresented sources of model errors, such as errors in the forcing and boundary conditions, is to use covariance inflation. The relaxation-to-prior spread approach (Whitaker and Hamill, 2012) was therefore adopted in this study. In this way, the posterior ensemble standard deviation is relaxed back to the prior using the following formula:

$$x'^a_i \leftarrow x'^a_i \left(\alpha\frac{\sigma^b - \sigma^a}{\sigma^a} + 1\right) \tag{3}$$

where $x'$ is the deviation from the posterior ensemble mean for the $i$th ensemble member, $\sigma^b$ and $\sigma^a$ are the prior and posterior ensemble standard deviations and $\alpha$ is the relaxation coefficient. We adopted a relaxation coefficient equal to 1.0, which means that the posterior ensemble standard deviation is completely replaced by the prior ensemble standard deviation, leaving an unaltered ensemble mean. Such an inflation is applied to the ensemble with the data assimilation. The prior ensemble standard deviation coincides with the standard deviation calculated from the free ensemble, while the posterior one is computed from the ensemble with the data assimilation after the forecast run following the analysis step. This type of inflation aims to reintroduce the ensemble spread necessary for the system to properly assimilate the observations while leaving the spatial correlations associated with the covariance matrix unaltered, since the ensemble mean is maintained.

#### 2.3.2. Covariance localization

The reliability of the EnKF technique primarily depends on an adequate representation of the error covariance that is maintained throughout the cycling of the data assimilation system. The generation of spurious covariances between distant or physically disconnected state vectors may lead to ensemble collapse and/or filter divergence. Covariance localization is therefore introduced to limit the generation of spurious covariances (Sakov and Bertino, 2011). The localization approach excludes remote observations from each analyzed location, thereby improving the conditioning of the error covariance matrices. Covariance localization in the EnKF is commonly implemented through a correlation matrix holding the correlations of local support defined by an influencing distance (see below). Other alternatives to the implementation of covariance localization inversely scale the observation error variance rather than modify the background error covariance (Hunt et al., 2007), replace the sample covariance with a multiscale tree of nodes distributed over a relatively small number of discrete scales (Zhou et al., 2008), filter the error covariance after a transformation to a spectral space (Buehner and Charron, 2007) or are raised to a power ensemble correlation (Bishop and Hodyss, 2009). Previous studies on geophysical flows (Houtekamer and Mitchell, 2001), biological modeling (Hu et al., 2012), and water resources (Devegowda et al., 2010) have shown that covariance localization substantially improves the EnKF analysis results.

The update equations are modified by replacing the state error covariance by its element-wise (Schur-Hadamard) product with a distance-based correlation matrix $\rho_l$:

$$K_i = [\rho_l \circ (P_i^f H_i^T)][\rho \circ (H_i P_i^f H_i^T) + R_i]^{-1} \tag{4}$$

Here, ρ is determined using a fifth-order piecewise rational function, as given by Gaspari and Cohn (1999):

$$\rho_c = \begin{cases} 1 - \frac{1}{4}\left(\frac{e}{l}\right)^5 + \frac{1}{2}\left(\frac{e}{l}\right)^4 + \frac{5}{8}\left(\frac{e}{l}\right)^3 - \frac{5}{3}\left(\frac{e}{l}\right)^2 & 0 \le e \le l \\ \frac{1}{12}\left(\frac{e}{l}\right)^5 - \frac{1}{2}\left(\frac{e}{l}\right)^4 + \frac{5}{8}\left(\frac{e}{l}\right)^3 + \frac{5}{3}\left(\frac{e}{l}\right)^2 - \frac{5}{3}\left(\frac{e}{l}\right) + 4 - \frac{2}{3}\left(\frac{e}{l}\right)^{-1} & l \le e \le 2l \\ 0 & e \ge 2l \end{cases} \quad (5)$$

where $e$ is the Euclidean distance between a grid point and an observation location and $l$ is the horizontal influence radius. Only observations located within a specified distance (defined by $l$) from an analyzed grid point will contribute to the analysis at that grid point. The empirical orthogonal function (EOF) technique (Hannachi et al., 2007) has been applied to the salinity and temperature datasets extracted from the 10-year reanalysis of the Copernicus Marine Environment Monitoring Service (CMEMS) catalog. The horizontal influence radius was calculated from the horizontal covariogram of the first EOF spatial mode, which relates the covariance between the locations of two measurements with their relative horizontal distances for a stationary process. Considering a 50% correlation threshold, the empirical covariogram analysis resulted in an average horizontal correlation length of 30 km.

Three model versions are operated every 48 h, providing forecasts over the subsequent two-day period. The first version is the ensemble free run, in which the ensemble members that are initialized as described above run independently without assimilating data. Each member is one-way nested in the MFS as indicated in Section 2.2. Seven forecast cycles are run from 7 to 21 August. The other two versions are ensembles that assimilate the glider observations collected during the 48 h preceding the analysis time both with and without the application of covariance localization. After the analysis step, a 48-hour forecast step is performed. Seven assimilation cycles are run while assimilating the observations collected from 5 to 19 August.

Finally, the observations are treated following Mourre and Chiggiato (2014). The salinity and temperature observations collected by the glider Jade are first interpolated onto a fixed model grid and stretched in the vertical direction before being assimilated. For each level, the observed variance in the vertical grid cell is used as an approximation of the vertical representation of the error. The horizontal representation of the error variance is assumed to be $(0.05)^2$ for the salinity field in the region (Mourre and Chiggiato, 2014).

### 2.4. Defining the model skill

Quantitatively, the prediction performance of the model is measured using statistical metrics that compare the model estimates or predictions with pairwise-matched observations (Willmott et al., 2012). Most statistical metrics can be encoded by a dimensionless index of an agreement or a skill score defined as $\rho = 1 - \delta / \mu$, where $\delta$ is a dimensional measure of the average model error and $\mu$ is the basis of comparison. The value of $\rho$ determines the forecast that performs the best. Different definitions of $\delta$ and $\mu$ generate different indices of agreement (see Willmott et al., 2012 for details). Here, $\delta$ and $\mu$ are the mean square error and the variance of the observations, respectively, averaged over the upper 300 m of the water column where the NC and the associated flow of the MAW are located. The forecast shows a positive (negative) skill $\rho$ when $\delta < \mu$ ($\delta > \mu$) and shows no skill ($\rho = 0$) when $\delta = \mu$. The forecast perfectly agrees with the observations ($\rho = 1$) when $\delta = 0$.

The mean square error and the forecast skill can be written in terms of the standard deviations of the model and observations ($\sigma_{m\,o}$, where $m$ and $o$ denote the model and the observations, respectively), the squared difference between those standard deviations (the square of the standard deviation error $SDE = \sigma_m - \sigma_o$), the cross-correlation (CC) between the modeled and observed fields, and the model bias ($MB = \overline{m} - \overline{o}$, where the overbars denote the mean values) (Murphy, 1992). The first term measures the statistical variability in the observations and in the corresponding predictions, while the second term quantifies the difference between both variabilities. The cross-correlation provides an indication of the similarity between the spatial patterns of the modeled and observed variables, while the last statistic measures how much the model is biased with respect to the observations. In order to complement the forecast skill, information concerning the above mentioned variables is also provided in an analysis of the model performance.

In this work, the forecast skills and associated statistical variables rely on the spatial averages of the forecasted and observed fields, thereby facilitating the tracking of the model's performance through time when a spatiotemporal series of observations is available. However, information about the similarity between modeled and observed spatial patterns is rather limited. Direct comparisons between the spatial distributions of the forecast and the observed fields at specific times were conducted to provide a qualitative spatial analysis that complements the quantitative assessment provided by the forecast skill.

## 3. Results

### 3.1. Predicted ocean circulation

This analysis describes the performances of different assimilation configurations for forecasting the path and velocity of the core of the NC, which corresponds to the most energetic part of the current. Thus, the horizontal current and vertically integrated salinity within the first 100 m of the water column were analyzed for the different model configurations. To serve as ground truth for the model forecast, the local circulation pattern was derived from along-ship tracking of the ocean current measurements for the period between 5 and 9 August 2013. The velocity field derived from the ADCP measurements from 5 to 9 August 2013 provides a rather detailed description of the circulation in the area, as shown in Fig. 2. The most evident feature is the NC, which is shown as an intense northwestward current entering from the southwestern boundary that flows across most of the study area before turning around the shelf break and then exiting the domain to the northwest. There is also an anticyclonic eddy centered at approximately 21 km from the coast with a radius of 16 km and a slightly elongated shape in the north-south direction. The resulting current speed averaged over the first 100 m of the water column is approximately 0.2 m $s^{-1}$. This horizontal circulation pattern provides a synoptic view of the conditions at the beginning of the experiment.

Unfortunately, the ADCP data collected after 9 August were not dense enough to provide a reliable synoptic environmental picture for the following validation periods. However, Borrione et al. (2016) indirectly inferred that the circulation pattern did not change significantly throughout the whole experimental period. This finding was also confirmed by the small temporal variability observed in the repetitive validation sections (see below).

The forecasts from the free run are shown in Fig. 3a, b and c for the beginning (9 August), middle (15 August) and end (19 August) of the experimentation period, respectively. The results show very little variability in the model outputs during the period under consideration. The NC is represented as a smooth and wide current that cyclonically crosses the region of interest. An anticyclonic eddy exists between the NC and the coast, but the eddy is positioned 36 km to the southeast from the observed location. The resulting vertically averaged current speed is 0.1 m $s^{-1}$. Regarding the vertically averaged salinity field over the first 100 m of the water column, the model initially predicted the existence of a water mass with a salinity of < 38 associated with the anticyclonic eddy (Fig. 3a). This water mass became saltier as the model evolved due to the mixing of water masses from the open sea, as seen in Figs. 3b and c.

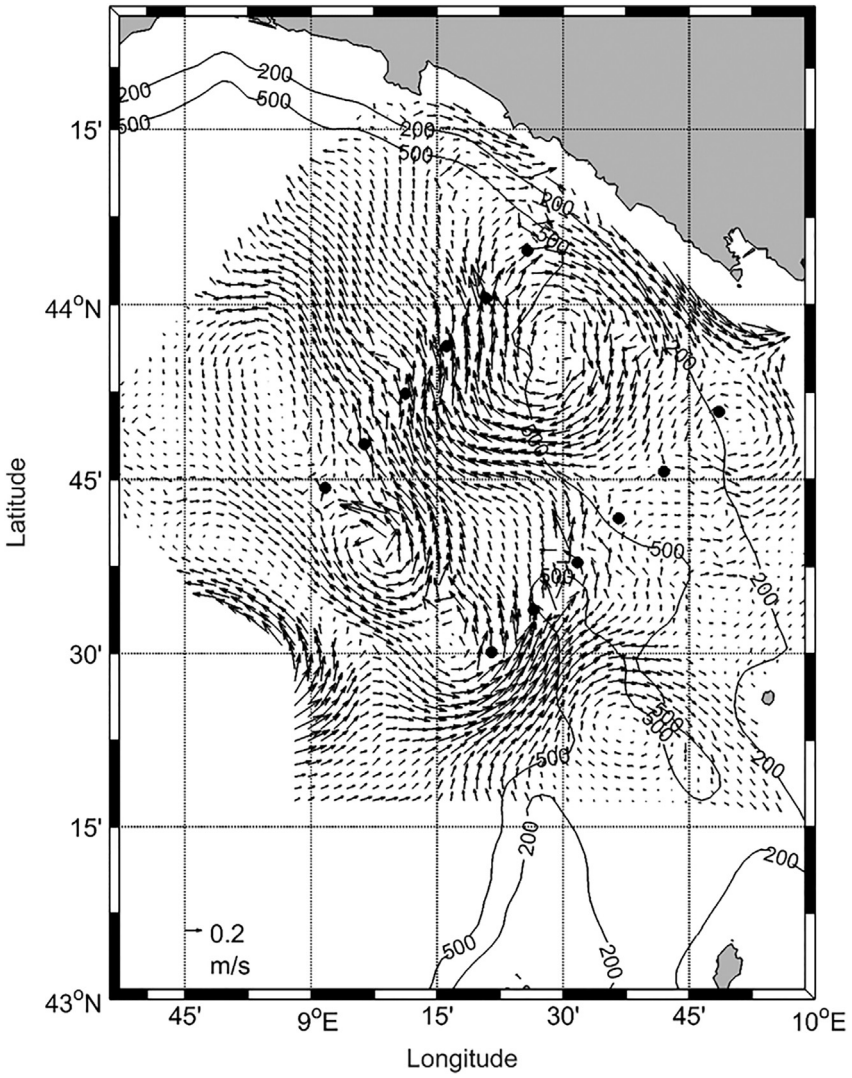


**Fig. 2.** Integrated horizontal current field from 0 to 100 m obtained from the ADCP data. The CTD locations are represented by black circles, while the light gray lines correspond to the isobaths of 200 and 500 m depth.

The sequence of the model forecast for the integrated horizontal current and salinity (with data assimilation but without covariance localization) is displayed in Fig. 3d (9 August), e (15 August) and f (19 August). During the first two assimilation cycles (Fig. 3d and e), the glider started sampling the coastal water masses moving towards the middle of the cyclonic gyre and later turned northwest to reach the northernmost region of the domain. Fig. 3d shows the circulation predicted by the model after the data collected along the described trajectory were assimilated. The alignment between the NC and the salinity front is evident from the figure. The jet crossed the northern and southern CTD transects nearly perpendicularly. The maximum integrated speeds are found along the middle (offshore) portions of the northern (southern) CTD transect. Note that the salinity gradient across the northern section is steeper than that across the southern region. A clear deficiency in this forecast is the lack of anticyclonic circulation between the NC and the coastline.

Fig. 3e shows the circulation pattern predicted by the model after the data collected along the glider trajectory from 9 to 15 August were assimilated. During this period, the glider was commanded to loop back in order to reach a location between the two CTD sections on 15 August. Except for the northernmost part of the trajectory, the glider sampled the MAW inside the NC cyclonic gyre between 9 and 11 August. The resulting width of the jet is reduced relative to that of the previous pattern (Fig. 3a), and the jet presents a more pronounced curvature. Accordingly, the model also predicted a shift of the 38.3 isohaline to follow the modified jet circulation (Fig. 3e). A strong readjustment of the 38.2 isohaline is consequently observed. A weakening of the salinity gradient across the CTD transects is detected from these modifications. Finally, a weak anticyclonic eddy is generated between the NC and the coast. However, the location of the eddy is 26 km to the south of the position where it was observed.

During the final assimilation cycle (Fig. 3f), the glider was directed towards the southern CTD transect on 16 August while following a trajectory that was almost perpendicular to the jet flow; thus, the glider sampled the fresher water located to the east of the southern CTD section from 17 to 19 August. Fig. 3f shows that the NC flow develops a strong positive curvature associated with local cyclonic circulation to the east of the southern section. In this cycle, salinity values higher than 38.5 are restricted to a very limited area. The predicted anticyclonic eddy remains weaker than the observed one, but its location becomes closer to its observed position due to a northwestward drift of approximately 16 km from its previous location.

An interesting feature is observed when analyzing the sequence of predictions of this model setup. Specifically, a comparison between Fig. 3d, e and f reveals that the spatial coherence among the current and salinity fields degrades as the data assimilation and model integration proceed. The initially smooth patterns of the current and salinity fields (Fig. 3d) evolve into structures with significantly smaller scales after the different assimilation cycles (Fig. 3f). This change could have been caused by the loss of spatial coherence in the covariance matrix each

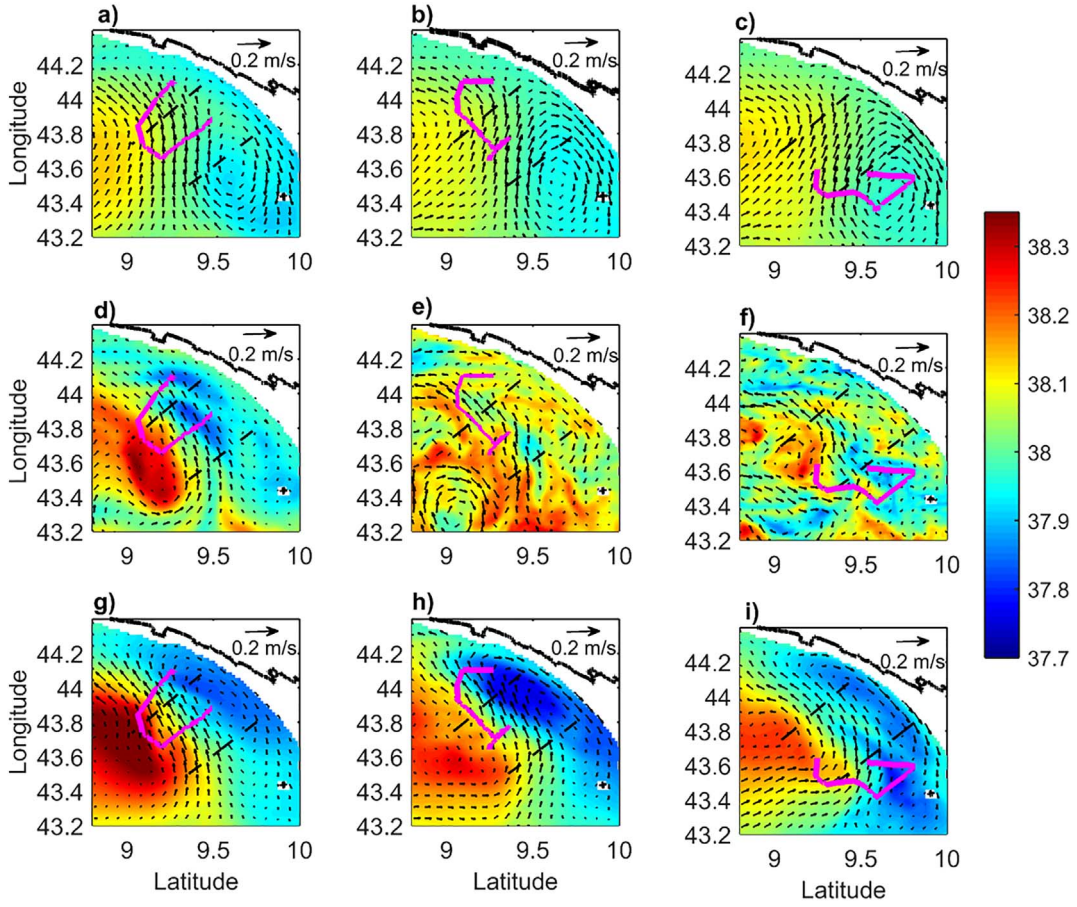


**Fig. 3.** Integrated horizontal current and salinity from 0 to 100 m for (a) 9 August, (b) 15 August and (c) 19 August obtained from the free run. Figures d), e) and f) are similar to a), b) and c) but for the model with data assimilation. Figures g), h) and i) refer to the model with data assimilation and covariance localization. The CTD transects are represented by the black dashed lines. The magenta lines indicate the horizontal glider track along which the data collected were assimilated to produce the forecast on 9, 15 and 19 August. Colour scale refers to salinity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

time that it is updated by the EnKF procedure. This hypothesis is confirmed by Fig. 4a and b, which show the correlation map for the salinity field at a depth of 50 m at an arbitrary location along the glider trajectory during the second and third assimilation cycles, respectively. These figures show a clear loss of spatial coherence in the correlation field with the ongoing assimilation cycles. After the assimilation cycles are terminated, the correlation field for the selected location is described by small-scale patches of correlated areas within the study domain. Thus, assimilation without covariance localization could generate an error covariance that correlates analyzed locations with distant, irregular, localized regions in the domain. This is expected to degrade the performance of the EnKF.

As in the previous cases, Fig. 3g, h and i display the forecasts obtained using the model setup with assimilation and covariance localization for 9 August, 15 August and 19 August, respectively. The NC flows along a well-defined salinity front between the coast and the open sea. The overall circulation of the current resembles the structure built from the observations. A weak anticyclonic circulation is present between the current and the coast. This anticyclonic circulation is reinforced in the forecast for 15 August due to the assimilation of glider data (Fig. 3h). Specifically, the salinities of the water masses between the current and the coast are lowered due to the assimilation of glider observations of low-salinity waters (~37.7) in this region. Nevertheless, the predicted anticyclonic circulation is more elongated than in the observations (Fig. 2). The salinity of the open seawater is also reduced with respect to the previous forecast (Fig. 3g). The lack of glider sampling in this region prevents the retention of relatively high salinity values in the open sea, as they are smoothed by advection and diffusion processes in the model. Note the different impacts of the assimilation of glider data when comparing this forecast (Fig. 3h) with the corresponding forecast in the case of assimilation without covariance localization, as seen in Fig. 3e. The data assimilated by the model comes from the transit of the glider across the southern part of the domain of interest. Salinity gradients in the anticyclonic eddy are not supported by glider observations, as they are smoothed out by advection and diffusion processes in the model. As a consequence, the anticyclonic eddy becomes weak and geographically less defined. Similar to the previous assimilation case, the impacts of covariance localization on the preservation of spatial coherence in the updated covariance matrix with ongoing cycling of the assimilation system are analyzed in Fig. 4c and d. These figures show that this reduction is significantly smaller in this case than in the case without covariance localization, although the spatial coherence is reduced (Fig. 4a and b). The assimilated data in this case evidently influence the modeling results at greater distances than in the assimilation setup without covariance localization, consequently producing well-defined ocean structures in the model forecast.

### 3.2. *Performance evaluation along the northern CTD transect*

The performance evaluations along the CTD transects will focus on the upper 300 m of the water column, which is where the NC and the associated flow of the MAW are located. Fig. 5a, b, c, and d characterize the model performance when the glider is located at near, intermediate and far distances from the CTD transect, respectively. Fig. 5a reveals the transit of the MAW across the transect during the experiment. The isohaline contours show clear temporal variability with a vertical

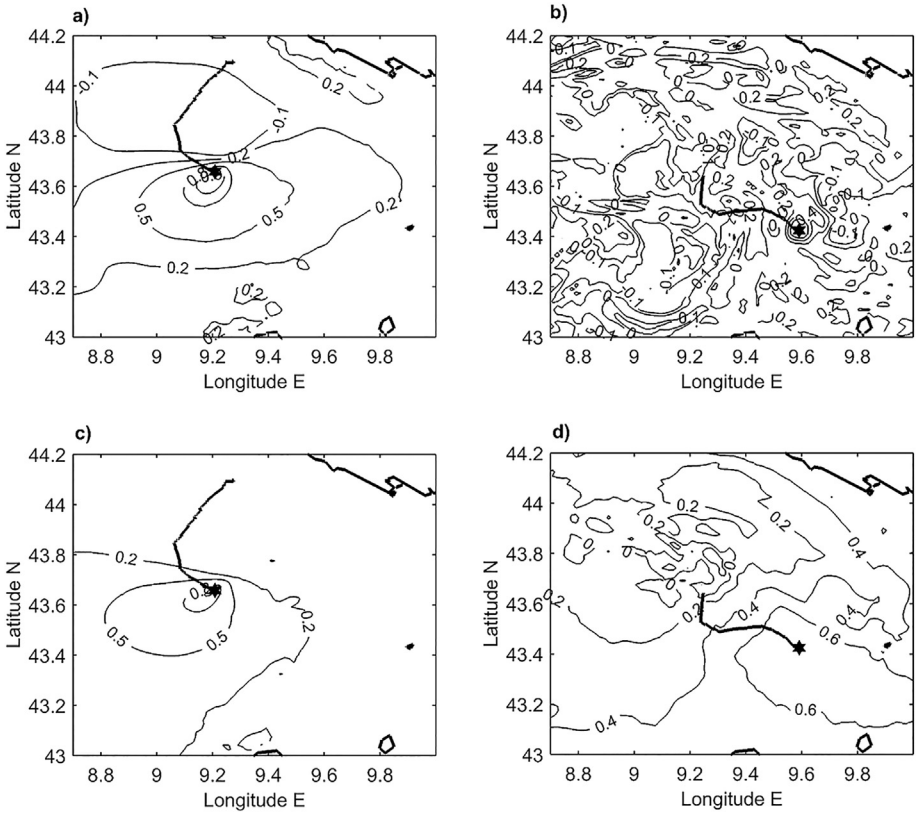


**Fig. 4.** Correlation map of the salinity at 50 m depth for the (a) second and (b) last assimilation cycles for the case without covariance localization. Contours display the correlation between the salinity at the location indicated by the black star and the corresponding values at the remaining grid points in the domain. Solid lines represent the glider track. Panels (c) and (d) refer to the second and last assimilation cycles for the case with covariance localization applied, respectively.

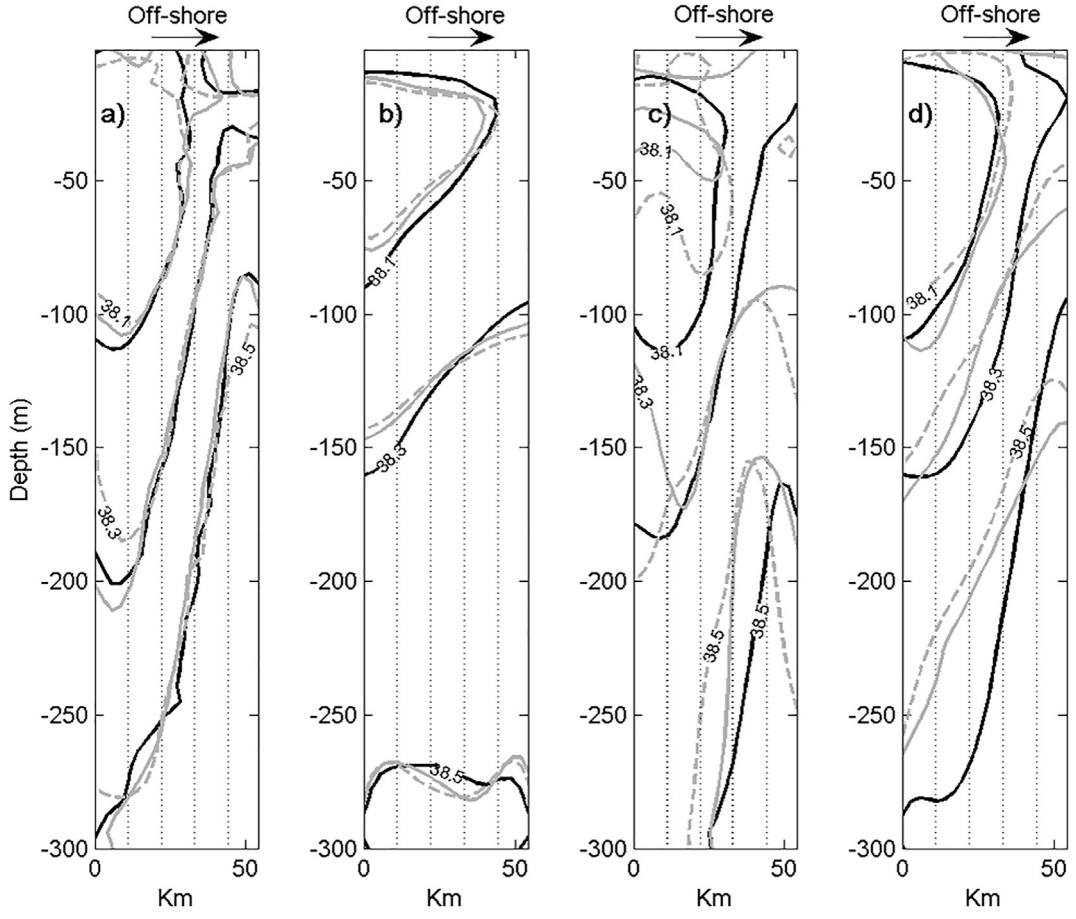


**Fig. 5.** CTD northern transect: Observed (a) and predicted salinity field for the case (b) without the assimilation of glider data, (c) with the assimilation but without covariance localization, and (d) with the assimilation and with covariance localization applied. The solid black, solid gray and dashed gray lines refer to 10, 16 and 19 August, respectively.

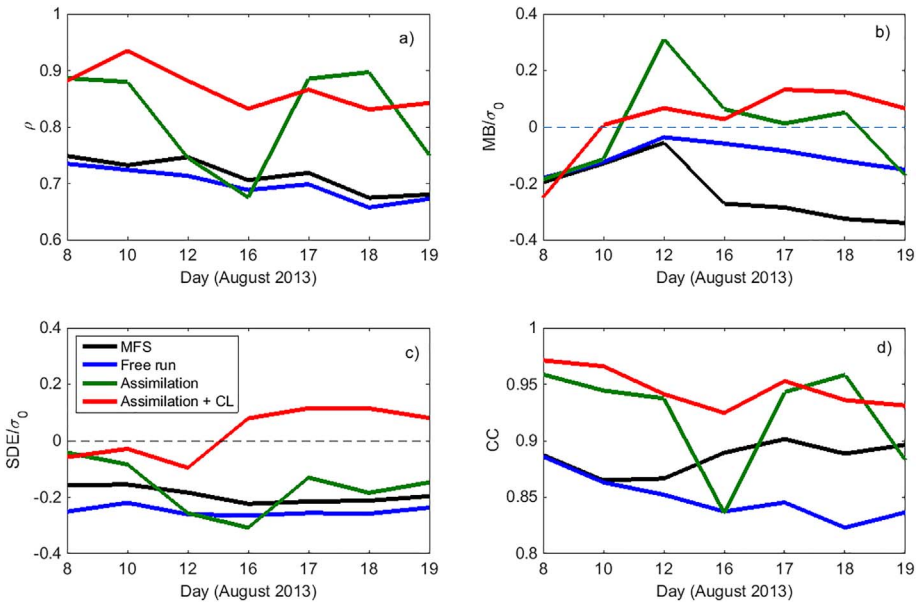


**Fig. 6.** Depth averaged (0–300 m) statistics for the salinity forecasts at the northern CTD transect showing (a) the skill index, (b) the normalized MB, (c) the normalized squared SDE and (d) the CC. The black, blue, green, and red lines refer to the MFS forecast, the free ensemble and the assimilation solution without and with covariance localization (CL), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contraction of > 50 m in the upper layer. The absence of a significant horizontal spread associated with this contraction seems to suggest that its origin is related to natural variability in the transport of the MAW rather than to dynamical effects (i.e., a gain in the relative vorticity). In particular, the northern section is located in the middle of the strong salinity front, which characterizes the NC cyclonic gyre (Fig. 2). The onshore CTD stations in the first 100 m of the water column are subjected to the transit of the MAW along the Ligurian coast with salinities ranging from 37.8 to 38.1. The offshore CTD stations are characterized by the transit of saltier MAW in the first 100 m of the water column with salinity values ranging from 38.1 to 38.4 that are bounded at the bottom by a more saline water mass (38.5). The most obvious temporal variability in the observed vertical pattern is associated with a shift in the MAW in the surface layers from a patched structure to a more compact shape. The observed vertical structures resemble that of a boundary current in the geostrophic balance flowing along an isobath.

The free ensemble run shows very limited variability during each forecast cycle (Fig. 5b). The saltier isohalines show a rather weak gradient in comparison with the observed ones; the saltier waters are confined to the intermediate water layers, while the observed 38.3 and 38.5 isohalines reach shallower depths of 50 m and 100 m, respectively.

The forecasts with data assimilation but without covariance localization improve upon the results of the free run during the first assimilation cycle, as shown in Fig. 5c. The horizontal location and depth of the core of the current system more closely resemble the observations (Fig. 5a) than the forecasted results from the free run (Fig. 5b). Deficiencies are also present, however, as is seen in the predicted depth of the 38.5 isohaline. The forecasts degrade with ongoing time and as the glider leaves the neighborhood of the validation transect. The salinity pattern resulting from the first assimilation cycle is significantly distorted in successive assimilation cycles as a result of internal model dynamics. These results indicate that the impact of data assimilation is not strong enough to constrain the model dynamics.

The benefits of data assimilation with covariance localization as the glider moves across the northern part of the domain are evident when comparing Fig. 5c, d and b with Fig. 5a.

The assimilation of glider data significantly corrects the spatial geometries of the isohalines predicted by the free ensemble by adjusting their corresponding depths and slopes to be more similar to the observations. As in the case without covariance localization, the performance of the forecast for the northern CTD transect is degraded as the glider starts to move towards the southern part of the study domain, as shown in Fig. 5d. The degradation of the performance is more evident for the deeper isohalines that are below the diving depth of the glider. This degradation becomes even more noticeable during the final assimilation cycle when the glider moves into the southern region far from the northern CTD transect. Despite the observed deficiencies, the forecasted temporal evolution of the haline structure along the northern CTD section obtained in this case is more similar to the observations than the case with assimilation but without covariance localization.

Figs. 6a, b, c and d display the temporal evolution of ρ and the MB, SDE and CC, respectively, for the forecasts from the different model configurations and the MFS. The skill score ρ changes from 0.73 to 0.68 during the free run experiments, as seen in Fig. 6a. The near constant value of ρ with time is probably associated with the reliability of the wind forecasts and the internal evolution of the model dynamics. The forecasts with assimilated data generally have higher ρ values than those obtained by the free run. However, there is an important difference between the skill indices from the model with assimilation only and that with assimilation and covariance localization. The former shows a strong irregular behavior over time with values bounded between the initial value of ρ and those produced by the free run. This finding is another indication that the performance of the forecast may not be robust to the location of the assimilated data. Conversely, the value of ρ decreases relatively smoothly from its initial value of 0.9 to the final value of 0.84 for the model with both data assimilation and covariance localization. Fig. 5d shows that the discrepancies between the model results and observations are mainly concentrated in the subsurface layers, and thus, the degradation of the performance may be due to the southward trajectory of the glider at the end of the experiment. Finally, note the similar magnitudes and evolution of ρ corresponding to the forecasts generated by the free run ensemble and the MFS.

The evolutions of the MB, SDE and CC (Figs. 6b, c and d) resemble the characteristics described above for ρ. In general, the model with assimilation performed better than the free run and MFS forecasts. The MFS forecast is affected by a negative model bias that grows from an initial value of $-0.2$ to a final value of $-0.34$. The free run does not show such a degradation of the MB, which registers a maximum value of $-0.18$. The evolution of the SDE shows a trend similar to those of the MFS forecast and free run forecasts, while the MFS forecast shows a higher correlation with the salinity field at the end of the experiment than the free run. The indices of the forecasts from the model with data assimilation and without covariance localization show a sharp and

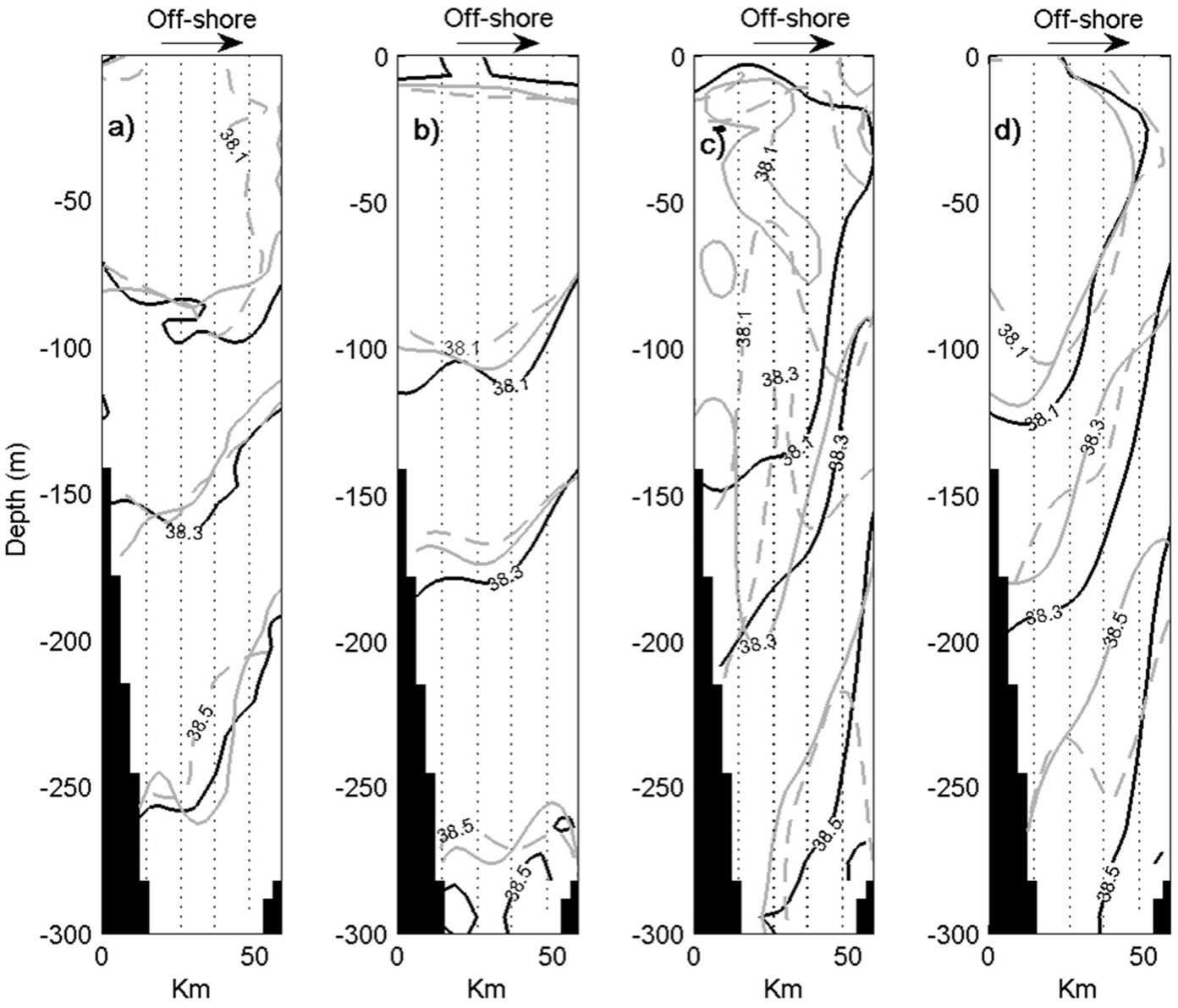


**Fig. 7.** Southern transect: Observed (a) and predicted salinity field for the case (b) without assimilation of glider data, (c) with assimilation but without covariance localization, and (d) assimilation with covariance localization applied. The solid black, solid gray and dashed gray lines refer to 9, 15 and 19 August, respectively. Black shaded areas represent the bathymetry.

irregular variability. Finally, the forecast from the model with both assimilation and covariance localization shows the best performance among the three indices (the MB, SDE and CC). Moreover, the temporal variability in this model is relatively smooth and shows a higher degree of robustness than the approach with only assimilation.

### *3.3. Performance evaluation along the southern CTD transect*

A performance analysis similar to that conducted for the northern CTD section is applied to the southern CTD transect. The southern section is located along the margin of the NC salinity front on its off-shore side, and it is not intersected by the sharp salinity gradients that characterize the northern section (Fig. 7a). The MAW flows across this section in the first 100 m of the water column with salinities $< 38.1$. An ill-defined vertical structure characterized by the existence of different patches with variable shapes in both space and time suggests that the observed variability can be understood in terms of the transit of diverse bifurcations of the local current system. Specifically, an analysis of the ADCP data (Fig. 2) acquired during the campaign reveals that this transect intersects the southern boundary of an anticyclonic eddy along its nearshore segment and intersects the inflow jet of the NC along its offshore segment.

Figs. 7b and c display the evolution of the model forecast from the free run and the forecast from the model with assimilation but without covariance localization, respectively, for 9 August, 15 August and 19 August. Qualitatively, no significant benefits from the assimilation of glider data are observed in the temporal evolution of the salinity field except for the deeper isohaline, which shows a trend in both the slope and depth similar to those observed on 15 August. For the shallowest isohaline, the forecasts from the model without assimilation produce better matches with the observations than the forecasts from the model that assimilates glider data throughout the entire forecasting period. Note that the doming of the isohalines predicted for 19 August by the model with assimilation indicates the generation of a cyclonic circulation in the model results, as shown in Fig. 7c. This predicted circulation was not observed in the period under consideration. The degradation of the forecast appears to be a direct consequence of the trajectory of the glider during the second and third assimilation cycles (Fig. 3f) and the

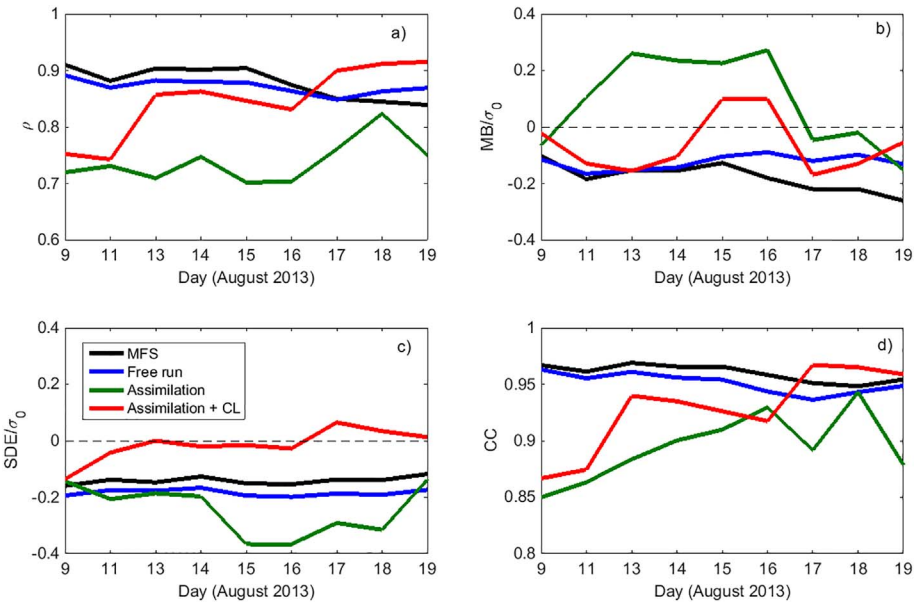


**Fig. 8.** Same as Fig. 6 but for the southern validation transect.

associated flow-dependent model error covariance. In general, we observe that the free ensemble reproduced the spatial pattern reasonably well, as the reproduced pattern is coherent with the large-scale dynamics of the Ligurian Sea; meanwhile, the assimilated ensemble was affected by spurious model anti-correlations at a high-frequency scale that propagate inside the model domain with the ongoing cycling of the data assimilation system. As explained in Section 2, an inflation of ensemble perturbations is employed here to prevent the collapse of the ensemble variance during the assimilation cycles. While the resulting standard deviation agrees with that of the free ensemble, the spatial correlation associated with the updated covariance matrix is unaltered. Finally, the forecast from the model configuration with data assimilation but without covariance localization shows a strong temporal variability that does not correspond to the observations.

The forecasts from the model with both assimilation and covariance localization are shown in Fig. 7d. The results still show deficiencies when compared to the observations. Specifically, the onshore isohalines are deeper than the observed ones. This difference from the observations is slowly corrected as the assimilation proceeds, at least with regard to the 38.1 and 38.5 isohalines. Note that the temporal variations in the forecasts are weaker than in the previous case and are more similar to the observations. This finding agrees with the forecast evaluation along the northern transect, where the use of covariance localization smooths out the sensitivity of the model forecast to the newly assimilated data.

The temporal evolutions of ρ and the MB, SDE and CC are displayed in Fig. 8a, b, c and d, respectively, for the southern transect. The free run forecasts show high values of ρ during the different assimilation cycles. Meanwhile, the models with assimilation initially generate oceanographic conditions in the southern region that are far from reality. Unlike the model with only data assimilation, the model with both assimilation and covariance localization shows rapid increases in the value of ρ as the glider approaches the southern region and terminates with the best index during the last assimilation cycle. Regarding the normalized MB and SDE, Fig. 8b and c show that all models generate forecasts with different non-zero values of the MB and SDE. The MB of the free run remains nearly constant throughout the simulation period, while the MFS forecast shows an increasingly negative bias after 15 August, ultimately reaching a maximum value of $-0.26$ at the end of the experiment. The MB values for the models with assimilation vary in form and intensity, as seen in Fig. 8b. Both models with data assimilation try to correct the negative model bias in the free run; the model without covariance localization shows a maximum MB of 0.26 against the lower value of 0.1 for the model with covariance localization. At the end of the experiment, both of the models attain a negative model bias; the model with covariance localization registers a lower model bias of $-0.05$. The correction of the SDE during the assimilation cycles of the model with both assimilation and covariance localization is the most remarkable fact that is displayed in Fig. 8c. Note that the initial non-zero value of the SDE is not corrected by only assimilating the data without covariance localization. Finally, the evolution of the CC for each different forecast is shown in Fig. 8d. The CC values for the free run and the MFS remain nearly unaltered with high correlation values for the duration of the experiment. However, the forecast patterns from the models with assimilation become more correlated with the observations as the assimilation proceeds. The predicted patterns obtained when the assimilation is complemented with covariance localization results in the forecast with the highest correlation with the observations. In summary, the figures seem to indicate that data assimilation when the glider is far from the validation region induces a degradation in the forecast quality with respect to the free run forecast. However, the initial deficiencies are corrected when the model assimilates data with covariance localization and when the glider observations are representative of the oceanographic conditions within the validation region.

### 3.4. *Assessment of the model biases*

The analysis of the MB values displayed in Fig. 6b and 8b deserves further attention. This is because the existence of an MB in the free run model may significantly degrade the performance of the assimilation scheme. MBs can be mainly introduced through open boundaries or atmospheric forcing, among other sources. An identification of the MB sources can be conducted from a comparison of long-term model runs with available observations. In contrast, due to the limited size of the open sea, long-term environmental knowledge is sparse or inexistent in the region considered. Alternatively, a comparison of the data gathered during the ScanFish transect along the periphery of the surveyed domain with the different model results could support an effort to discern whether MBs are introduced through the open boundaries. Fig. 9a displays the salinity field sampled along the ScanFish track (see Fig. 1), while Fig. 9b, c and d show the model errors obtained from the MFS, the free run, the model with data assimilation but without covariance localization and the model with both assimilation and covariance

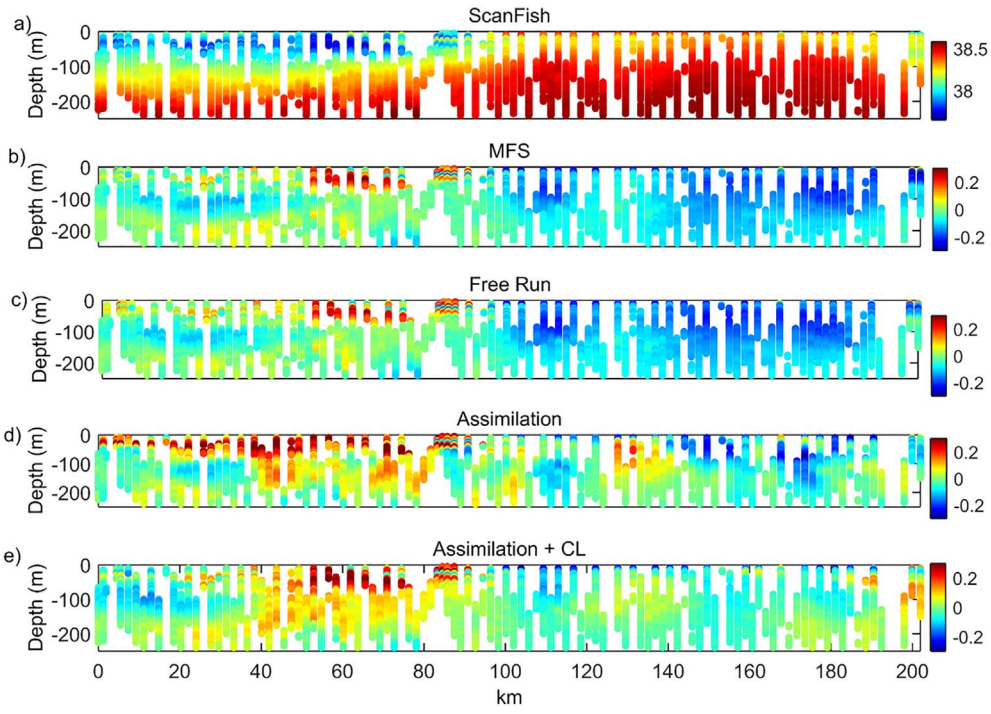


**Fig. 9.** a) Salinity observations obtained with the ScanFish MKII vehicle along the periphery of the region considered. Salinity differences between observations and b) MFS, c) free run, d) model with assimilation without covariance localization and e) model with assimilation and covariance localization.

localization, respectively. A comparison of Fig. 9a and b reveals that the MFS results underestimate the salinity along the sampled track. This is in agreement with the negative values of the MFS MB reported in Figs. 6b and 8b. Moreover, the magnitude and spatial pattern of the salinity errors obtained from the free run (Fig. 9c) closely resemble those from the MFS (Fig. 9b). This supports the hypothesis that MBs are mostly introduced through the open boundaries. The similar initial evolutions of the MBs associated with the MFS and the free run (Figs. 6b and 8b) also reinforce this hypothesis. Fig. 9d and e provide insight about the robustness of the different assimilation procedures to the existence of biases. The results from the model run with both data assimilation and covariance localization significantly correct the salinity differences observed from the other runs (Fig. 9e). Again, this result is in agreement with the evolution of the MB displayed in Figs. 6b and 8b.

## 4. Discussion and conclusions

The availability of near-daily CTD casts at fixed spatial locations has allowed for a detailed temporal and spatial validation of the performance of a regional ROMS-EnKF system used to assimilate temperature and salinity observations collected by a Slocum glider over a 15-day period. The performance assessment was conducted by tracking the temporal evolution of a commonly used performance index and by comparing forecast patterns with observations along two predetermined transects located halfway between the centerline of the domain of interest and its northern and southern boundaries. The latter was possible thanks to the sustained collection of independent observations.

A set of numerical simulations were run to consider the production of near-operational forecasts for the region through a free run ensemble, a model configuration including data assimilation, and a model configuration that augmented the data assimilation with covariance localization. It was found that the forecast of the salinity field in the northern validation section using the model with data assimilation generally improves the ρ, MB, SDE and CC estimates relative to the free run. The impact of assimilation was favorable for the simple circulation structure in that portion of the study domain characterized by a geostrophically adjusted current associated with a salinity front that evolves at a slow scale. Thus, a glider transect inside the salinity front is sufficient for constraining the model state to a dynamical evolution closer to the observations from the first assimilation cycle. For this reason, data assimilation corrects the wide structure of the NC derived from the free run. The performance indices obtained from the model with both assimilation and covariance localization improved upon those obtained from the other cases considered. Moreover, they were also more robust to the location of the assimilated observations. Data assimilation was not very efficient at improving the model forecasts in the southern validation section. The complicated spatial circulation pattern along the southern transect with a current bifurcation and recombination clearly reduces the ability of the model ensembles to produce reliable forecasts in this region. Assimilating data when the glider transits the northern part of the domain significantly degrades the performances of the model predictions in the southern area with respect to the free run. This initial difference in the performance is reduced and overcome by using the assimilation approach incorporating covariance localization when the glider reaches the southern validation transect.

As was previously mentioned (see Fig. 4a and b and the related text in Subsection 3.1), the estimation of the covariance in the EnKF degrades essential properties, leading to the underestimation of the analysis error and the generation of spurious long-range correlations. The results presented here also indicate the loss of spatial coherence in the covariance, which is another undesirable deficiency introduced during the update process. These detrimental effects start after the second assimilation cycle and propagate over the subsequent successive cycles, resulting in model anti-correlations at a high spatial frequency that become meaningful when the glider moves outside of the NC gyre. It is hypothesized that the generation of the spurious sample covariance matrix originates from the introduction of higher-frequency time scales (uncorrelated with the slow scales) in the system that eventually dominate the first sample covariance matrix derived from the free run ensemble. These fast, uncorrelated scales progressively degrade the EnKF performance and, in particular, attenuate and distort the salinity front. This degradation mechanism resembles the mechanism identified by Ballabrera-Poy et al. (2009) in a test case for a nonlinear chaotic model coupled on two different spatiotemporal scales. These authors found a detrimental effect in the EnKF when introducing fast, evolving scales uncorrelated with the rest of the dynamic system. This generates

a spurious sample covariance matrix when using a finite ensemble. Fast, evolving scales are produced during the adjustment of the analysis field in the present case. The sorting of the assimilation periods dictated by the fast and local dynamics prevents an adequate computation of the numerical covariance.

The positive impacts of covariance localization are more evident for the northern validation section, where the oceanographic structure of the water column is described by a geostrophic current flowing along the continental slope. Moreover, the distance of the glider from this section generates a slower degradation of the forecast performance. However, in the southern validation transect, such positive effects are observed only when the glider is in relatively close proximity to the transect. The dynamical complexity of the circulation pattern in the southern region of the study domain increases the sensitivity of the forecast performance to the sampling strategy. Although some studies have attempted to address the definition of the most effective sampling strategy based on the available resources and the dynamical variability of the region (Alvarez and Mourre, 2012; Mourre and Alvarez, 2012; Heaney et al., 2016), this issue still remains largely unexplored.

Our findings are aligned with previous works that analyzed the benefits of the EnKF assimilation scheme together with covariance localization in regional simulations. Keppenne and Rienecker (2003) identified multiple key factors, including an inhomogeneous, anisotropic background error covariance model, a dynamical estimation of the covariance model and a possibility for multivariate updates, that can improve the performance of the EnKF when compared with a standard univariate optimal interpolation. A similar conclusion was reached by Miyazawa et al. (2012) when studying the interactions among the open and coastal seas. The need for an EnKF approach is justified in terms of the highly variable scales originating from the coexistence of multi-scale phenomena, the statistical representation of which requires an anisotropic and continuously updated error covariance. Covariance localization is considered a fundamental methodology for filtering long-range correlations that result from small ensemble sizes employed to estimate the anisotropic error covariance (Miyazawa et al., 2012). Our results complement previous studies by highlighting another beneficial mechanism resulting from covariance localization. Specifically, it limits the detrimental effects originating from the introduction of fast, evolving scales that might be uncorrelated with the rest of the dynamic system. To this end, covariance localization was identified as an efficient way for partially mitigating these spurious effects. This approach is especially relevant in the framework of this experiment, where short-term forecasts (i.e., 2 days) were considered. This time scale limited the possibility of applying other approaches for filtering out undesirable high frequencies.

Data assimilation schemes using the EnKF and covariance localization are not exempt from further improvement. An important issue to consider is the fact that model errors are not unbiased. In the present study, the model bias was derived from a large-scale model and was then introduced to the regional model through the open boundaries. The presence of model biases may generate additional difficulties when observations are localized in a relatively small region of the domain. A bias can produce field discontinuities in modeling areas far away from the observations. An approach that can be used to mitigate the effects of error biases consists of the assimilation of different data types and the use of observations with global coverage, if available, even when studying spatially limited areas (Keppenne and Rienecker, 2003). In this regard, satellite observations are particularly useful. However, the assimilation of satellite data was not specifically considered in our work. Rather than providing an operational forecast, the purpose of this study was to isolate and identify the mechanisms that degrade the estimation of the error covariance in the EnKF as well as to evaluate the efficiency of covariance localization in mitigating those errors when modeling a limited coastal area. This approach could be hindered if these mechanisms are masked by the assimilation of a rich set of observations. Moreover, limitations in satellite data coverage and/or quality during an experiment could prevent their exploitation.

To conclude, this study has shown that the use of covariance localization can significantly improve the performance of an EnKF for a field experiment with a dense data set that was continuously collected to allow for a unique model forecast validation. This improvement both prevents the occurrence of spurious long-range correlations in the covariance and contributes to preserving the spatial coherence in the covariance matrix throughout the estimation process in the EnKF.

## Acknowledgments

We acknowledge the Italian Air Force National Meteorological Center for the provision of the COSMO-ME data. We thank Dr. Baptiste Mourre for providing the coupled EnKF ROMS code used in this study. We are grateful to the CMRE staff for maintaining and piloting the gliders, for the CTD instruments and to the Alliance crew for the operations conducted on the NRV Alliance. This work was funded by the Alliance Command Transformation (ACT project number SAC000507) and the EU project FP7 (284321)-INFRASTRUCTURES-2011-1 “Gliders for research, ocean observation and management-GROOM”.

## References

Alvarez, A., Mourre, B., 2012. Optimum sampling designs for a glider-mooring observing network. J. Atmos. Ocean. Technol. 29, 601–612. http://dx.doi.org/10.1175/JTECH-D-11-00105.1.

Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimiliation. Mon. Weather Rev. 129, 2884–2903. http://dx.doi.org/10.1175/15200493(2001)129<2884:AEAKFF>2.0.CO;2.

Anderson, J.L., Anderson, S.L., 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon. Weather Rev. 126, 2741–2758. http://dx.doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.

Ballabrera-Poy, J., Kalnay, E., Yang, S.-C., 2009. Data assimilation in a system with two scales-combining two initialization techniques. Tellus 61A, 539–549. http://dx.doi.org/10.1111/j.1600-0870.2009.00400.x.

Birol, F., Cancet, M., Estournel, C., 2010. Aspects of the seasonal variability of the Northern current (NW Mediterranean Sea) observed by altimetry. J. Mar. Syst. 81, 297–311. http://dx.doi.org/10.1016/j.jmarsys.2010.01.005.

Bishop, C.H., Hodyss, D., 2009. Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models. Tellus A 61, 84–96. http://dx.doi.org/10.1111/j.1600-0870.2008.00371.x.

Borrione, I., Falchetti, S., Alvarez, A., 2016. Physical and dynamical characteristics of a 300 m-deep anticyclonic eddy in the Ligurian Sea (Northwest Mediterranean Sea): evidence from a multiplatform sampling strategy. Deep-Sea Res. I 116, 145–164. http://dx.doi.org/10.1016/j.dsr.2016.07.013.

Brasseur, P., 2006. Ocean data assimilation using sequential methods based on the Kalman filter. In: Chassignet, E., Verron, J. (Eds.), Ocean Weather Forecasting SE. 10. Springer, Netherlands, pp. 271–316. http://dx.doi.org/10.1007/1-4020-4028-8_10.

Buehner, M., Charron, M., 2007. Spectral and spatial localization of background-error correlations for data assimilation. Q. J. R. Meteorol. Soc. 133, 615–630. http://dx.doi.org/10.1002/qj.50.

Chao, Y., Li, Z., Farrara, J.D., Moline, M.A., Schofield, O.M.E., Majumdar, S.J., 2008. Synergistic applications of autonomous underwater vehicles and the regional ocean modeling system in coastal ocean forecasting. Limnol. Oceanogr. 53, 2251–2263. http://dx.doi.org/10.4319/lo.2008.53.5_part_2.2251.

Chapman, D.C., 1985. Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model. J. Phys. Oceanogr. 15, 1060–1075. http://dx.doi.org/10.1175/1520-0485(1985)015<1060:NTOCSO>2.0.CO;2.

Devegowda, D., Arroyo-Negrete, E., Datta-Gupta, A., 2010. Flow relevant covariance localization during dynamic data assimilation using EnKF. Adv. Water Resour. 33 (2), 129–145. http://dx.doi.org/10.1016/j.advwatres.2009.10.001.

Dobricic, S., Pinardi, N., Testor, P., Send, U., 2010. Impact of data assimilation of glider observations in the Ionian Sea (Eastern Mediterranean). Dyn. Atmos. Oceans 50, 78–92. http://dx.doi.org/10.1016/j.dynatmoce.2010.01.001.

Edwards, C.A., Moore, A.M., Hoteit, I., Cornuelle, B.D., 2015. Regional Ocean Data Assimilation. Annu. Rev. Mar. Sci. 7, 21–42. http://dx.doi.org/10.1146/annurev-marine-010814-015821.

Evensen, G., 1994. Inverse methods and data assimilation in nonlinear ocean models. Physica D 77, 108–129. http://dx.doi.org/10.1016/0167-2789(94)90130-9.

Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. 53, 343–367. http://dx.doi.org/10.1007/s10236-003-0036-9.

Falchetti, S., Alvarez, A., Onken, R., 2015. A Relocatable EnKF Ocean Data Assimilation Tool for Heterogeneous Observational Networks. IEEE/MTS OCEANS 2015 – Genovapp. 1–6. http://dx.doi.org/10.1109/OCEANS-Genova.2015.7271359.

Flather, R.A., 1976. A tidal model of the northwest European continental shelf. Mem. Soc. R. Sci. Liege 6, 141–164. http://dx.doi.org/10.1175/JTECH1930.1.

Gangopadhyay, A., Schmidt, A., Agel, L., Schofield, O., Clark, J., 2013. Multiscale

forecasting in the western North Atlantic: sensitivity of model forecast skill to glider data assimilation. Cont. Shelf Res. 63, 1–18. http://dx.doi.org/10.1016/j.csr.2012.09.013.

Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. Q. J. R. Meteorol. Soc. 125, 723–757. http://dx.doi.org/10.1002/qj.49712555417.

Haidvogel, D.B., Arango, H., Budgell, W.P., Cornuelle, B.D., Curchitser, E., Di Lorenzo, E., Fennel, K., Geyer, W.R., Hermann, A.J., Lanerolle, L., Levin, J., McWilliams, J.C., Miller, A.J., Moore, A.M., Powell, T.M., Shchepetkin, A.F., Sherwood, C.R., Signell, R.P., Warner, J.C., Wilkin, J., 2008. Ocean forecasting in terrain-following coordinates: formulation and skill assessment of the Regional Ocean Modeling System. J. Comput. Phys. 227, 3595–3624. http://dx.doi.org/10.1016/j.jcp.2007.06.016.

Hannachi, A., Jolliffe, I.T., Stephenson, D.B., 2007. Empirical orthogonal functions and related techniques in atmospheric science: a review. Int. J. Climatol. 27, 1119–1152. http://dx.doi.org/10.1002/joc.1499.

Heaney, K.D., Lermusiaux, F.J., Duda, T.F., Halley, P.J., 2016. Validation of genetic algorithm based optimal sampling for ocean data assimilation. Ocean Dyn. 66, 1209–1229. http://dx.doi.org/10.1007/s10236-016-0976-5.

Houtekamer, P.L., Mitchell, H.L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129, 123–137. http://dx.doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., Hansen, B., 2005. Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. Mon. Weather Rev. 133, 604–620. http://dx.doi.org/10.1175/MWR-2864.1.

Hu, J., Fennel, K., Martten, J.P., Wilkin, J., 2012. Data assimilation with a local ensemble Kalman filter applied to a three-dimensional biological model of the Middle Atlantic Bight. J. Mar. Syst. 94, 145–156. https https://doi.org/10.1016/j.jmarsys.2011.11.016.

Hunt, B.R., Kostelich, E.J., Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. Phys. D 230, 112–126. http://dx.doi.org/10.1016/j.physd.2006.11.008.

Jones, E.M., Oke, P.R., Rizwi, F., Murray, L.M., 2012. Assimilation of glider and mooring data into a coastal ocean model. Ocean Model 47, 1–13. http://dx.doi.org/10.1016/j.ocemod.2011.12.009.

Kantha, L.H., Clayson, C.A., 1994. An improved mixed layer model for geophysical applications. J. Geophys. Res. 99, 25235–25266. http://dx.doi.org/10.1029/94JC02257.

Keppenne, C.L., Rienecker, M., 2003. Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. J. Mar. Sys. 40–41, 363–380. http://dx.doi.org/10.1016/S0924-7963(03)00025-3.

Keppenne, C.L., Rienecker, M.M., Kurkowski, N.P., Adamec, D.A., 2005. Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction. Nonlinear Process. Geophys. 12, 491–503. http://dx.doi.org/10.5194/npg-12-491-2005.

Lenartz, F., Mourre, B., Barth, A., Beckers, J.M., Vandenbulcke, L., Rixen, M., 2010. Enhanced ocean temperature forecast skills through 3-D super-ensemble multi-model fusion. Geophys. Res. Lett. 37, 1–5. http://dx.doi.org/10.1029/2010GL04459.

Lermusiaux, P.F.J., 2007. Adaptive modeling, adaptive data assimilation and adaptive sampling. Physica D 230, 172–196. http://dx.doi.org/10.1016/j.physd.2007.02.014.

Lermusiaux, P.F.J., Chiu, C., Gawarkiewicz, G.G., Abbot, P., Robinson, A.R., Miller, R.N., Haley, P.J., Leslie, W.G., Majumdar, S.J., Pang, A., Lekien, F., 2006. Quantifying uncertainties in ocean predictions. Oceanography 19 (1), 90–103. http://dx.doi.org/10.5670/oceanog.2006.93.

Lorenz, E.N., 1969. The predictability of a flow which possesses many scales of motion. Tellus 21, 289–307. http://dx.doi.org/10.1111/j.2153-3490.1969.tb00444.x.

Marullo, S., Salusti, E., Viola, A., 1985. Observations of a small-scale baroclinic eddy in the Ligurian Sea. Deep Sea Res. Part A 32, 215–222. http://dx.doi.org/10.1016/0198-0149(85)90029-9.

Millot, C., 1999. Circulation in the Western Mediterranean Sea. J. Mar. Syst. 20, 423–442. http://dx.doi.org/10.1016/S0924-7963(98)00078-5.

Miyazawa, Y., Miyama, T., Varlamov, S.M., Guo, X., Waseda, T., 2012. Open and coastal seas interactions south of Japan represented by an ensemble Kalman Filter. Ocean Dyn. 62, 645–659. http://dx.doi.org/10.1007/s10236-011-0516-2.

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

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

Murphy, A.H., 1992. Climatology, persistence, and their linear combination as standards of reference in skill scores. Weather Forecast. 7, 692–698.

Pinardi, N., Coppini, G., 2010. Preface "Operational oceanography in the Mediterranean Sea: the second stage of development". Ocean Sci. 6, 263–267. http://dx.doi.org/10.5194/os-6-263-2010.

Ramp, S.R., Davis, R.E., Leonard, N.E., Shulman, I., Chao, Y., Robinson, A.R., Marsden, J., Lermusiaux, P.F.J., Fratantoni, D.M., Paduan, J.D., Chavez, F.P., Bahr, F.L., Liang, S., Leslie, W., Li, Z., February 2009. Preparing to predict: the Second Autonomous Ocean Sampling Network (AOSN-II) experiment in the Monterey Bay. Deep-Sea Res. II Top. Stud. Oceanogr. 56 (3–5), 68–86. ISSN 0967-0645. https://doi.org/10.1016/j.dsr2.2008.08.013.

Robinson, A., Sellschopp, J., 2002. Rapid assessment of the coastal ocean environment. In: Pinardi, N., Woods, J. (Eds.), Ocean Forecasting SE. 11. Springer, Berlin Heidelberg, pp. 199–229. http://dx.doi.org/10.1007/978-3-662-22648-3_11.

Sakov, P., Bertino, L., 2011. Relation between two common localization methods for the EnKF. Comput. Geosci. 15, 225–237. http://dx.doi.org/10.1007/s10596-010-9202-6.

Schroeder, K., Ribotti, A., Borghini, M., Sorgente, R., Perilli, A., Gasparini, G.P., 2008. An extensive western Mediterranean deep water renewal between 2004 and 2006. Geophys. Res. Lett. 35, 1–7. http://dx.doi.org/10.1029/2008GL035146.

Shulman, I., Rowley, C., Anderson, S., DeRada, S., Kindle, J., Martin, P., Doyle, J., Cummings, J., Ramp, S., Chavez, F., Fratantoni, D., Davis, R., 2009. Impact of glider data assimilation on the Monterey Bay model. Deep. Res. Part II Top. Stud. Oceanogr. 56, 188–198. http://dx.doi.org/10.1016/j.dsr2.2008.08.003.

Turner, M.R.J., Walker, J.P., Oke, P.R., 2008. Ensemble member generation for sequential data assimilation. Remote Sens. Environ. 112, 1421–1433. http://dx.doi.org/10.1016/j.rse.2007.02.042.

Umlauf, L., Burchard, H., 2003. A generic length-scale equation for geophysical turbulence models. J. Mar. Res. 61, 235–265. http://dx.doi.org/10.1357/002224003322005087.

Unesco, 1985. The International System of Units (SI) in Oceanography. UNESCO Technical Papers No. 45, IAPSO Pub. Sci. No. 32, Paris, France.

Webb, D.C., Simonetti, P.J., Jones, C.P., 2001. SLOCUM: An underwater glider propelled by environmental energy. IEEE J. Ocean. Eng. 26, 447–452. http://dx.doi.org/10.1109/48.972077.

Whitaker, J.S., Hamill, T.M., 2012. Evaluating methods to account for system errors in ensemble data assimilation. Mon. Weather Rev. 140, 3078–3089. http://dx.doi.org/10.1175/MWR-D-11-00276.1.

Whitaker, J.S., Hamill, T.M., Wei, X., Song, Y., Toth, Z., 2008. Ensemble data assimilation with the NCEP global forecast system. Mon. Weather Rev. 136, 463–482. http://dx.doi.org/10.1175/2007MWR2018.1.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. Oceans Atmos. 90, 8995–9005. http://dx.doi.org/10.1029/JC090iC05p08995.

Willmott, C.J., Robeson, S.M., Matsuura, K., 2012. A refined index of model performance. Int. J. Climatol. 32, 2088–2094. http://dx.doi.org/10.1002/joc.2419.

Zhang, S., Harrison, M.J., Rosati, A., Wittenberg, A., 2007. System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. Mon. Weather Rev. 135, 3541–3564. http://dx.doi.org/10.1175/MWR3466.1.

Zhang, W.G., Wilkin, J.L., Arango, H.G., 2010a. Towards an integrated observation and modeling system in the New York Bight using variational methods. Part I: 4DVAR data assimilation. Ocean Model 35, 119–133. http://dx.doi.org/10.1016/j.ocemod.2010.08.003.

Zhang, W.G., Wilkin, J.L., Levin, J.C., 2010b. Towards an integrated observation and modeling system in the New York Bight using variational methods. Part II: Representer-based observing strategy evaluation. Ocean Model 35, 134–145. http://dx.doi.org/10.1016/j.ocemod.2010.06.00.

Zhou, Y., McLaughlin, D., Entekhabi, E., Crystal, G.H., 2008. An emsemble multiscale filter for large nonlinear data assimilation problems. Mon. Weather Rev. 136, 678–698. http://dx.doi.org/10.1175/2007MWR2064.1.

# Document Data Sheet

| Security Classification | | Project No. |
|---|---|---|
| | | |
| **Document Serial No.** | **Date of Issue** | **Total Pages** |
| CMRE-PR-2019-002 | May 2019 | 14 pp. |


**Author(s)**

Silvia Falchetti, Alberto Alvarez


**Title**

The impact of covariance localization on the performance of an ocean EnKF system assimilating glider data in the Ligurian Sea


**Abstract**

Data assimilation through an ensemble Kalman filter (EnKF) is not exempt from deficiencies, including the generation of long-range unphysical correlations that degrade its performance. The covariance localization technique has been proposed and used in previous research to mitigate this effect. However, an evaluation of its performance is usually hindered by the sparseness and unsustained collection of independent observations.

This article assesses the performance of an ocean prediction system composed of a multivariate EnKF coupled with a regional configuration of the Regional Ocean Model System (ROMS) with a covariance localization solution and data assimilation from an ocean glider that operated over a limited region of the Ligurian Sea. Simultaneous with the operation of the forecast system, a high-quality data set was repeatedly collected with a CTD sensor, i.e., every day during the period from 5 to 20 August 2013 (approximately 4 to 5 times the synoptic time scale of the area), located on board the NR/V Alliance for model validation. Comparisons between the validation data set and the forecasts provide evidence that the performance of the prediction system with covariance localization is superior to that observed using only EnKF assimilation without localization or using a free run ensemble. Furthermore, it is shown that covariance localization also increases the robustness of the model to the location of the assimilated data. Our analysis reveals that improvements are detected with regard to not only preventing the occurrence of spurious correlations but also preserving the spatial coherence in the updated covariance matrix. Covariance localization has been shown to be relevant in operational frameworks where short-term forecasts (on the order of days) are required.



**Keywords**

Data assimilation, Ensemble Kalman filter, Predictability, Regional ocean predictions, Underwater gliders, Ocean observing networks



**Issuing Organization**

NATO Science and Technology Organization
Centre for Maritime Research and Experimentation

Viale San Bartolomeo 400, 19126 La Spezia, Italy

*[From N. America:*
STO CMRE
*Unit 31318, Box 19, APO AE 09613-1318]*

Tel: +39 0187 527 361
Fax:+39 0187 527 700

E-mail: library@cmre.nato.int