# 4D-Var Data Assimilation and Observation Impact on Surface Transport of HF-Radar Derived Surface Currents in the North-Western Mediterranean Sea

Michele Bendoni[a], Andrew M. Moore[b], Anne Molcard[c], Marcello Magaldi[d], Maria Fattorini[a,e], Carlo Brandini[a,e]

[a]*LaMMA Consortium, 50019 Sesto Fiorentino (Fi), Italy*
[b]*Department of Ocean Sciences, University of California, 95062 Santa Cruz CA*
[c]*MIO, Université de Toulon, Aix-Marseille Université, CNRS, IRD, Toulon, France*
[d]*CNR-ISMAR, 19032 Lerici (Sp), Italy*
[e]*CNR-ISMAR, 50019 Sesto Fiorentino (Fi), Italy*

**Abstract**

We investigated the effect of the assimilation of surface velocities from two HF-Radars (HFRs) and satellite sea surface temperature (SST) into an ocean circulation model, implemented on the North-Western Mediterranean Sea, using the ROMS 4D-Var Data Assimilation (DA) System. A three months period from September to November 2020 is analysed. A model free run is used as reference simulation to quantify the impact of the data assimilation procedure on the surface circulation. The DA framework consists of a sequence of 3-day windows during which the observations are assimilated into the model, followed by forecast simulations, starting from the last time step of the analysis run. The comparison of simulated velocities and SST against the assimilated observations shows that both forecast and analysis runs improve the solution as compared to the freerun, with error reduction up to the 47% and doubling of correlation values. If SST is not assimilated, the simulated sea surface temperature slightly degrades as a result of the DA procedure. Furthermore, we used surface velocities from Lagrangian drifters to test the model against independent observations. The results show that both the forecast and the analysis simulate surface circulation better than the free run within the areas covered by HFR observations, with

*Michele Bendoni

error reduction ranging from 10% to 50%, and increase in correlation from 5% to more than doubled. Nevertheless, farther afield, outside the area covered by HFRs, improvements and degradations of the solution due to the DA procedure balance each other. The impact of assimilated observation on the alongshore transport of the upper 50 m is analysed along three transects adjacent to the two HFR areas and along a section in between. The transport increment is significant where the surface velocities are assimilated whereas it is negligible elsewhere. The effect of SST is to modify the velocity distribution along the transects, while keeping the transport increment unchanged. Modifications to the initial conditions have the most significant effect on the alongshore transport, whereas the corrections to boundary conditions and atmospheric forcing have different importance according to the analysed transects. The assimilation of surface velocities and SST affects more the ageostrophic component of the surface velocity field, rather than the geostrophic one.

## 1. Introduction

The dynamics of the ocean surface influence many important socio-economic aspects of the marine environment such as the transport of pollutants and marine litter (Cucco and Daniel, 2016; Fossi et al., 2017; Soto-Navarro et al., 2020; Van Sebille et al., 2020) and the analysis of hydrodynamic connectivity (Rossi et al., 2014; Ser-Giacomi et al., 2015), economic issues related to the transport of larvae and nursery areas (Quattrocchi et al., 2019), up to the safety of the marine environment with respect to search and rescue activities (Bellomo et al., 2015; Reyes et al., 2020).

The combined use of coastal HF-Radar systems (HFR) and ocean modelling tools allow us to greatly improve our knowledge of surface dynamics, especially if these two sources of information are combined using data assimilation techniques (DA). Early attempts to assimilate HFR data into coastal models are described by Breivik and Saetra (2001) and Paduan and Shulman (2004) using Optimal Interpolation and the physical-space statistical analysis system

(PSAS), respectively. Since then, significant progress has been made and HFR data are routinely assimilated from multi-platform systems in different areas of the world, especially through the use of 4D-Var DA. The 4D-Var technique has been employed in regional models in various parts of the world ocean, including along both the U.S. West coast (Yu et al., 2012; Moore et al., 2013) and East coast (Zhang et al., 2010), in the South Pacific (Janeković et al., 2013; Partridge et al., 2019), on the coasts of Africa (Phillipson and Toumi, 2017; Couvelard et al., 2021), Australia (Zavala-Garay et al., 2012), and the Norwegian Sea (Sperrevik et al., 2015).

Our focus here will be on the Mediterranean Sea which is served by several HFRs (Lorente et al., 2022), especially in the North-West area. This source surface current data has been used for various applications, including the ingestion into numerical models using different assimilation techniques (Reyes et al., 2022). Marmain et al. (2014) corrected atmospheric forcing and open boundary condition of a model of the French coast, through an Ensemble Perturbation Smoother. They found that by assimilating radial observations, the correction of the wind field affects surface currents more than the correction to the open boundary, which tends to operate over a deeper part of the water mass. Vandenbulcke et al. (2017) employed a deterministic Ensemble Kalman Filter (EnKF) to correct inertial oscillations in the Ligurian Sea. Hernandez-Lasheras et al. (2021) applied a local multi-model Ensemble Optimal Interpolation (EnOI) to improve the marine dynamics along the West coast of Spain. Iermano et al. (2016) used the ROMS 4D-Var DA system (Moore et al., 2011c) in a coastal model of the Gulf of Naples to assimilate surface currents from HFRs, sea surface temperature (SST) and sea surface height (SSH), and analysed the impact of single observations on modification of the alongshore transport. For the Adriatic Sea a multi platform system ingesting even HFR data has been developed by Janeković et al. (2020).

The main objective of the present study is to assess the impact of assimilating observations from two separate radar systems, together with satellite data of SST, into a model of the North-Western Mediterranean Sea using a 4D-Var

DA method. The two land-based radar systems are located on the French Mediterranean coast in the Toulon area (WERA system, Figure 1a,b), and on the Ligurian Italian coast in the La Spezia area (CODAR system, Figure 1a,c), 450 km apart, and enable to monitor the surface circulation over the radars' footprint (order of thousands $km^2$).

The North-Western Mediterranean Sea is characterized by a strong geostrophic and baroclinic boundary current (the Northern Current, NC) as part of the cyclonic general circulation of the basin. Associated with its seasonal variability, mesoscale to sub-mesoscale structures may appear along the flow and the positioning of the radars enables observation of the circulation upstream and at the core of the NC.

Section 2 describes the modelling setup, the 4D-Var DA framework and the observations employed. Section 3 presents a comparison of the modelled surface circulation with and without assimilation and the observation impact analysis. Last two sections are dedicated to the discussion of the results and the the main findings of the work, respectively.

## 2. Data Assimilation System

### *2.1. Model Configuration and Experimental Setup*

The Regional Oceanic Modelling System (ROMS) is a primitive equation, hydrostatic, terrain-following coordinate model (Shchepetkin and McWilliams, 2003, 2005), which is provided with a suite of tools to perform a 4D-Var DA (Moore et al., 2011c,a), observation impact and sensitivity analysis (Moore et al., 2011b).

For the present application the ROMS model is configured for an area of the North-Western Mediterranean Sea (NWM), spanning 2.40 E to 12.64 E and 41.60 N to 44.51 N (Figure 1). The model resolution is 2.3 km ($\sim 1/36^{\circ}$) in the horizontal and has 40 terrain following sigma-layers in the vertical. Turbulent horizontal eddy viscosity $\nu$ was set equal to 10.0 $m^2/s$ with harmonic horizontal diffusion of momentum along $\sigma$-surfaces, whereas eddy diffusivity coefficient

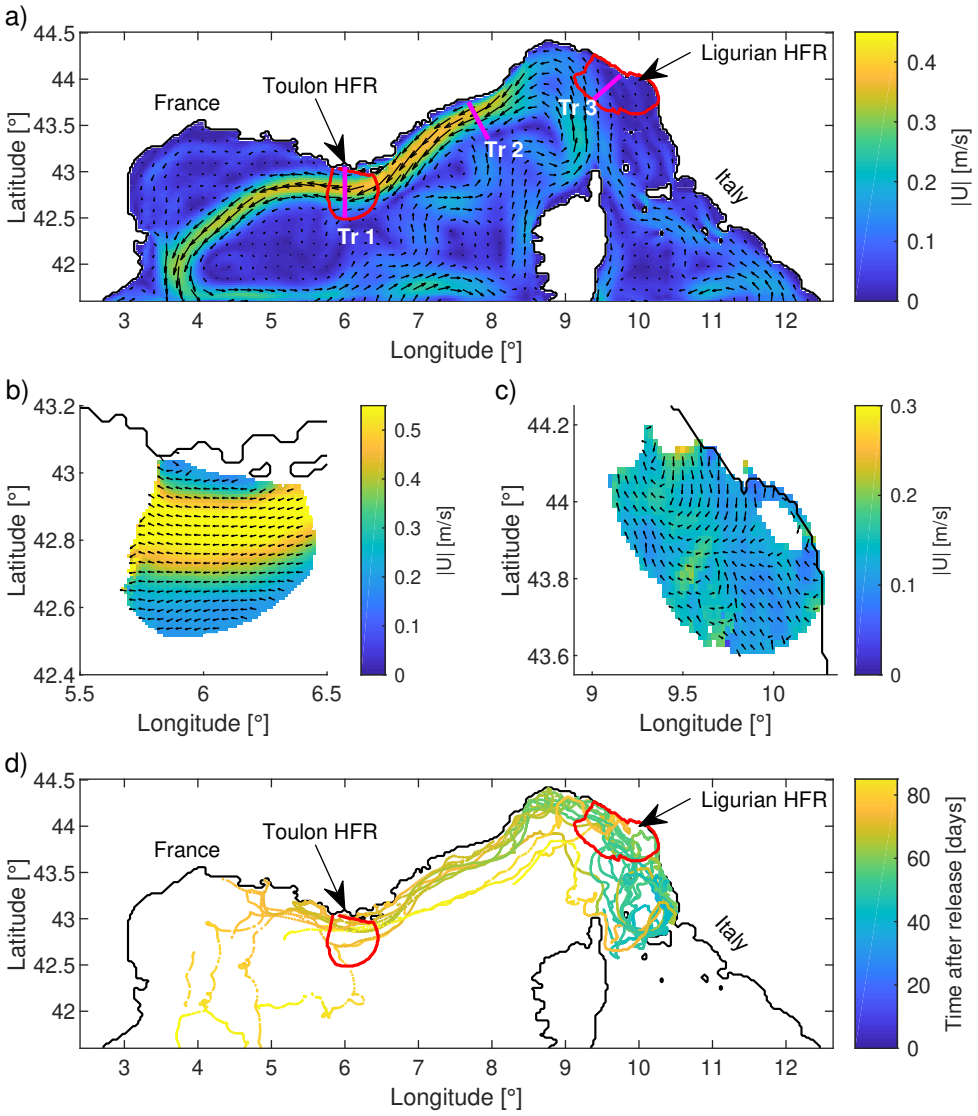


Figure 1: a) View of the extension of the ROMS domain and upper 50 m modelled current vector field superimposed to the velocity magnitude averaged over the period 1 September 2020 - 30 November 2020 for the free run. Red lines represent the extent of the radar data coverage for both the Toulon (TLN) and Ligurian (LIG) systems. The transects along which observation impacts are calculated are reported in magenta. b) View of the mean flow field and footprint measured by the TLN HFR. c) View of the mean flow field and footprint measured by the LIG HFR. d) Drifters trajectories as a function of the time elapsed since release in the Ligurian Sea.

$\kappa_T = 2$ m$^2$/s for both temperature and salinity with harmonic diffusion along geopotential surfaces. The $k-kl$ parametrization was used for the two-equation Generic Length Scale turbulence closure scheme (Warner et al., 2005). Air-sea fluxes were computed using bulk formulation of Fairall et al. (1996). Bathymetric data from the EMODnet database (http://www.emodnet-bathymetry-eu) were interpolated at the grid points and smoothed through a sequence of Laplacian filtering (Sikirić et al., 2009) to obtain $rx_0 = 0.2$ (Haidvogel et al., 2000) and $rx_1 = 7$ (Haney, 1991). Tidal forcing is not included in the modelling system since the amplitude of the tide is of the order of few centimetres in the analysed area.

The initial condition and the boundary conditions for the whole modelled period for surface height $\eta$, temperature $T$, salinity $S$ and horizontal velocities $u$ and $v$, at daily frequency with 1/24° resolution, come from the MEDSEA_ANALYSISFORECAST_PHY_006_013 product (Clementi et al., 2021), delivered by the Copernicus Marine Environment Monitoring Service (CMEMS). The surface velocities at hourly time scale coming from the same dataset are used for comparison with surface velocities modelled by the ROMS model. Clamped boundary conditions were applied to both velocity and tracers, a Flather condition (Flather, 1976) was imposed on the barotropic velocities, a Chapman condition (Chapman, 1985) to the free surface $\eta$, and zero gradient conditions to the total kinetic energy (TKE).

Atmospheric forcing was every 3-hours from the 1/10° High-Resolution (HRES) Atmospheric Model from ECMWF (https://www.ecmwf.int/en/forecasts/datasets/set-i). Specifically, the following variables are employed: air pressure at sea level, air temperature, air relative humidity, rainfall rate, surface winds and downward short- and long-wave radiations. River inputs for the Rhone and Arno rivers were based on daily averaged measurements of flow rates provided by the Hydro Portal of France Waters (https://www.hydro.eaufrance.fr/), and the Tuscany Region SIR Dataset (https://www.sir.toscana.it/), respectively.

The assimilation of observations into the model was performed using a 4D-Var approach. If $\mathbf{z}$ denotes the vector of control variables, then 4D-Var identifies

the model state which minimizes a nonlinear cost function defined as (Courtier et al., 1994):

$$J_{NL}(\mathbf{z}) = \frac{1}{2}[\mathbf{z} - \mathbf{z_b}]^T \mathbf{D}^{-1}[\mathbf{z} - \mathbf{z_b}] + \frac{1}{2}[\mathbf{y} - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y} - H(\mathbf{x})] \quad (1)$$

where $\mathbf{z}$ and $\mathbf{z_b}$ are the updated and background control vector, respectively, $\mathbf{D}$ is the background error covariance matrix, $\mathbf{R}$ is the observation error covariance matrix, $\mathbf{y} - H(\mathbf{x})$ is the innovation (usually identified with $\mathbf{d}$), $\mathbf{y}$ is the observation vector, $H$ is the observation operator, and $H(\mathbf{x})$ the observation operator applied to the state vector $\mathbf{x}$. The $\mathbf{D}$ matrix is a block diagonal matrix with blocks $\mathbf{B}_x$, $\mathbf{B}_b$ and $\mathbf{B}_f$, which are the background error covariance matrix for initial condition, boundary conditions and atmospheric forcing, respectively. In the application described here, the control vector comprises the state variables for the initial conditions, the open boundary conditions, and the surface fluxes of momentum, heat and freshwater.

To minimize $J_{NL}$, the incremental approach proposed by (Courtier et al., 1994) is adopted in the ROMS 4D-Var system (Moore et al., 2011c). This implies the linearization of the problem around the solution $\mathbf{z}_k = \mathbf{z}_b + \sum_{s=1}^{k-1} \delta \mathbf{z}_s$, where $k$ refers to the outer loop number and $\delta \mathbf{z}_s$ are the increments from the previous outer loops. Each outer loop corresponds to the minimization of the quadratic cost function (solved through a sequence of the so called inner loops):

$$J(\delta \mathbf{z}_k) = \frac{1}{2}\delta \mathbf{z}_k^T \mathbf{D}^{-1} \delta \mathbf{z}_k + \frac{1}{2}(\mathbf{G}_{k-1}\delta \mathbf{z}_k - \mathbf{d}_{k-1})^T \mathbf{R}^{-1}(\mathbf{G}_{k-1}\delta \mathbf{z}_k - \mathbf{d}_{k-1}) \quad (2)$$

where $\delta \mathbf{z}_k = \mathbf{z}_k - \mathbf{z}_{k-1}$ is the increment of the control vector associated to the $k^{th}$ outer loop and $\mathbf{G}_{k-1}$ is the linearized observation operator that samples the tangent linear model (linearised around the $(k-1)^{th}$ outer loop) at observation points.

The ROMS 4D-Var system supports several different 4D-Var approaches for minimizing the cost function by adjusting the control vector (Moore et al., 2011c). Here, we employ a dual formulation, in which the minimization was performed in the space spanned by the observations (Courtier, 1997). In addition, strong constraint formulation was used in conjunction with the Restricted

B-preconditioned Lanczos algorithm (Gürol et al., 2014) for the solution of the sequence of quadratic problems. Corrections were computed for the initial condition (IC) and boundary conditions (BC) for the prognostic variables sea surface height $\eta$, temperature $T$, salinity $S$ and horizontal velocities $u$ and $v$. Furthermore, the atmospheric forcing (AF), namely, the momentum surface stresses along $u$ and $v$ directions, and the heat and freshwater fluxes at the sea surface were also corrected by the 4D-Var algorithm.

For the experiments reported here, we first assimilated total surface currents estimated from two HFRs. Next, satellite SST observations were assimilated in addition to surface velocities. A detailed description of the assimilated data is reported in Subsection 2.2.

To evaluate the influence of DA on the model state estimates, we performed three sets of simulations: (i) a free-run (FR) where the model is run for 3 months, from 1 September to 31 November 2020, starting from a 15 days spin up period, without assimilation; (ii) a sequence of 30 cycles, each with a 3-day assimilation window, called the analysis (AN); and (iii) a series of model runs starting from the final time of each analysis window, called forecast (FC). The AN and FC runs were performed for the two combination of observations assimilated: HFR derived surface velocities alone, and surface velocities together with SST. Furthermore, surface velocities from the CMEMS product MEDSEA_ANALYSISFORECAST_PHY_006_013 (Clementi et al., 2021), computed with the NEMO model (NM) (Gurvan et al., 2022), are used as additional benchmark solution.

For the minimization of the cost function, we impose 2 outer loops and 7 inner loops, similar to Janeković et al. (2020) and Levin et al. (2021). This choice was based on a series of trials for the experiment with the assimilation of HFR data only, for which this combination gave the best results in terms of agreement between model and observations (not shown).

Background error covariances are simulated in ROMS 4D-Var following the diffusion operator approach (Weaver and Courtier, 2001), assuming that the background variances are proportional to the variability of the non-linear model

solution without assimilation. While ROMS 4D-Var supports a multivariate balance operator, previous experience has shown that the results obtained are somewhat mixed, and performance is generally better if the balance operator is disabled, and instead we rely on the adjoint model to establish the appropriate balances. As a consequence, the background error covariance matrix is as follows:

$$\mathbf{B} = \mathbf{\Sigma C \Sigma} \tag{3}$$

where $\mathbf{\Sigma}$ is the diagonal matrix of the standard deviations, and $\mathbf{C}$ is a symmetric matrix of background error correlations whose application on a vector can be expressed as the solution of a diffusion equation. More details about the modelling of the background covariance matrix can be found in Moore et al. (2011c) and Weaver and Courtier (2001). The standard deviations for the initial and boundary conditions and atmospheric forcing were estimated from the simulation of the three months window 1 September - 30 November for the years 2019, 2020 and 2021. For temperature and heat flux, the standard deviation was determined after each three months run was detrended and low-pass filtered with a 24-hours moving average to remove the effect of the diurnal cycle. The horizontal decorrelation length scales of the background errors were estimated from a semi-variogram analysis (Banerjee et al., 2003; Levin et al., 2021) applied to the same model runs used to estimate the background standard deviations. For the state variables, the following horizontal length scales were identified: 30 km for $\eta$, $T$ and $S$, and 15 km for $u$ and $v$. A vertical decorrelation length scale of 20 m was used for all the variables. This choice is related to the mixed layer depth that in the analysed area ranges between 10 and 40 m in the period September-November (Houpert et al., 2015). Moreover, a first trial with 30 m gave worse results. An error decorrelation length scale of 50 km was chosen for the atmospheric forcing fields.

The averaged velocity vector field and magnitude for the upper 50 m depth, over the three months period, is shown in Figure 1a. The main feature to note is the clear presence of the NC which flows adjacent to the French coastline with

an average velocity up to 0.45 m/s. Furthermore, it is possible to distinguish the Eastern and Western Corsica Currents (ECC and WCC), which merge north of Corsica, feeding the NC (Millot, 1999).

### *2.2. Observations*

Two sets of observations are assimilated into the model comprising surface velocities derived from HFR and sea surface temperature retrieved from satellite measurements. A third set of observations, consisting of surface velocities derived from surface Lagrangian drifters, has been used as independent dataset to further validate the DA procedure.

Surface velocities come from two different HFRs (Figure 1), one covering the area off Toulon (France), and the other one the area adjacent to La Spezia and Cinque Terre (Italy). The Toulon HFR (TLN) is a WERA system (Gurgel et al., 1999), with a resolution of 1 km and covers an area of roughly 55 km along-shore by 60 km cross-shore (Dumas et al., 2020). The TLN system has been installed since 2012 and has undergone numerous improvements since then, resulting in novel multi-static system (Guérin et al., 2019) and a widely validated long time series data (Bourg and Molcard, 2021; Molcard et al., 2021). Surface velocities from the Ligurian HFR (LIG) are retrieved from the INSITU_MED_NRT_OBSERVATIONS_013_035 CMEMS product. The HFR is a CODAR system (Barrick and Lipa, 1985), with 2 km resolution, and the area covered extends approximately 100 km in the along-shore direction and 45 km in the cross-shore direction. The monthly-averaged surface current field (November 2020) superimposed to the velocity magnitude $|\mathbf{U}|$, from both the HFR data is shown in Figures 1b and 1c.

Before assimilation, the HFR data were low-pass filtered to remove frequencies corresponding to periods shorter than 3 hours. Only those data with quality flag equal to "good" and $GDOP \leq 2$ were admitted to the assimilation system (Fang et al., 2015). As an additional filtering mechanism, we tried to remove data showing unphysical convergence and divergence patterns. First, the $u$ and $v$ velocity components were interpolated on a staggered grid, then, the values of

$|u_x + v_y|$ larger than $99^{th}$ percentile, within a 3 day assimilation window, were removed. Observations errors for the HFR data were set to a level of 0.1 m/s, similarly to Hernandez-Lasheras et al. (2021).

SST were obtained from the SST_MED_SST_L3S_NRT_OBSERVATIONS_010_012_b CMEMS product (Nardelli et al., 2013) at daily scale with information on temporal distribution, and 0.01° spatial resolution. The temperature error for a single observations was set to a level of 0.4° C, similarly to Moore et al. (2011a) and Iermano et al. (2016).

To reduce data redundancy, observations located in the same cell of the model domain and within hourly intervals, were averaged to form "super observations" (Moore et al., 2011a). The error variance associated to the $i^{th}$ super observation $\varepsilon_i$ was determined such that $\varepsilon_i^2 = \overline{\varepsilon_{o,i}}^2 + \sigma_{y,i}^2$, where $\overline{\varepsilon_{o,i}}$ corresponds to the average error and $\sigma_{y,i}$ to the standard deviation of observation values contained within the same grid cell within the same time interval, set equal to 1 hour. The firs term accounts for the instrument error, whereas the second one for the errors due to the interpolation of model values at observation locations and to the limitations of the model in representing all processes measured by the observation.

Surface drifters data come from the field measurement campaign DDR20- "Drifter demonstration and Research 2020" Berta et al. (2021). On October 8th 2020, 20 CARTHE-type biodegradable drifters (Novelli et al., 2017) were released in front of Livorno (Italy), on a regular grid of 3 km per side and 500 m pitch, within 2 hours. The drifters span the first 60 cm of water column and are specifically designed not to be affected by wind drag. Drifter trajectories were manipulated to obtain hourly data of surface $u_d$ and $v_d$ velocities. First, the geographical coordinates are transformed into metric coordinates through the M_Map package (Pawlowicz, 2020), the $u$ and $v$ velocity components are determined via central difference, and assigned to the mid point between subsequent positions. Then, geographical coordinates and velocities are hourly averaged.

*2.3. Observation Impact*

The ROMS 4D-Var system provides the tools to determine the contribution of each observation to a index $I(\mathbf{x})$, that describes some characteristics of the circulation, arising from the assimilation procedure. More specifically, the index increment $\Delta I = I(\mathbf{x}_a) - I(\mathbf{x}_b)$ can be approximated as follows (Langland and Baker, 2004; Moore et al., 2017; Levin et al., 2021):

$$\Delta I \simeq \mathbf{d}^T \mathbf{K}^T \mathbf{M}_b^T(t) \left. \frac{\partial I}{\partial \mathbf{x}} \right|_{\mathbf{x}_b} \tag{4}$$

where $\mathbf{K} = \mathbf{D}\mathbf{G}^T(\mathbf{G}\mathbf{D}\mathbf{G}^T + \mathbf{R})^{-1}$ is the Kalman gain matrix and $\mathbf{M}_b$ is the tangent linearization of the nonlinear model around the background circulation $\mathbf{x}_b$. Since the value of $\Delta I$ is computed as the dot-product of the innovation vector $\mathbf{d}^T$ and the vector $\mathbf{g} = \mathbf{K}^T \mathbf{M}_b^T(t) \partial I / \partial \mathbf{x}|_{\mathbf{x}_b}$, the contribution of the $i^{th}$ observation to the index increment $\Delta I$ is simply given by $d_i g_i$. In practice, $\mathbf{K}$ is replaced by its reduced rank approximation $\widetilde{\mathbf{K}}_m$ using the Lanczos vectors that result from the $m$ inner loops employed in the minimization of the cost function. A more detailed and thorough explanation of the procedure can be found in Moore et al. (2017) and Levin et al. (2021).

Since we employ two outer loops to minimize the cost function, the observation impact should be determined separately for each outer loop. In such a case, the impact due to the first outer loop $\Delta I_1$ is determined through Equation 4, whereas for the impact due to the second outer loop $\Delta I_2$, the innovation is $\mathbf{d} = \mathbf{y} - H(\mathbf{x}_a^1)$, and the tangent linearization, the reduced rank Kalman gain matrix and the derivative $\partial I / \partial \mathbf{x}$, are referred to the $\mathbf{x}_a^1$ analysis solution. However, following Trémolet (2008), and as shown by Levin et al. (2020), the total $\Delta I$ is for the most part ascribable to the first outer loop. Furthermore, the computation of $\Delta I_2$ is based on $\mathbf{x}_a^1$, which is, in turn, dependent on the observations. As a consequence, the contribution due to each observation is not unambiguously identifiable. For both reasons, from now on, observation impact will be referred only to the first outer loop.

Here, the chosen indeces are the alongshore mass transports across three transects up to 50 m depth: one close to the TLN HFR (Tr1), the second near

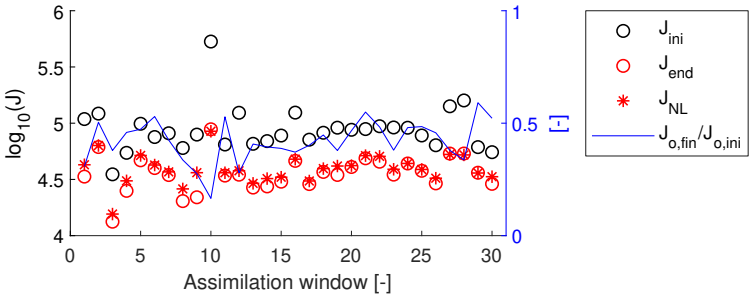


Figure 2: Evolution of the cost function $J$ as a function of the assimilation windows. Left vertical axis: values of the initial (black circles) and final (red circles) values of the cost function $J$ associated to the incremental form and the final value of the nonlinear cost function (red asterisks). Right vertical axis: ratio of the final to the initial value of the cost function associated to observations (blue line).

the France-Italy border between the two HFR systems (Tr2), and the third near the LIG HFR (Tr3) (Figure 1a). The alongshore transport is averaged within each assimilation window and it is considered positive in the direction of the NC.

## 3. Results

### *3.1. Performance of the DA System*

To evaluate the degree to which the assimilation procedure is able to fit the model to the observations, we report in Figure 2 the values of the initial and final linearized cost functions $J_{ini}$ and $J_{end}$ (Equation 2), together with the final value of the nonlinear cost function $J_{NL}$ (Equation 1). Also shown in Figure 2 is the ratio of the final to the initial cost function associated to observations $J_{o,fin}/J_{o,ini}$. Figure 2 shows the case for the assimilation experiment where both surface velocities and SST are assimilated into the system.

There is a clear reduction of $J$, and the observation component of the cost function $J_o$ is generally reduced by 50%. The values of $J_{end}$ and $J_{NL}$ are similar during the majority of cycles, indicating that the tangent linear assumption is substantially valid throughout the whole experiment.

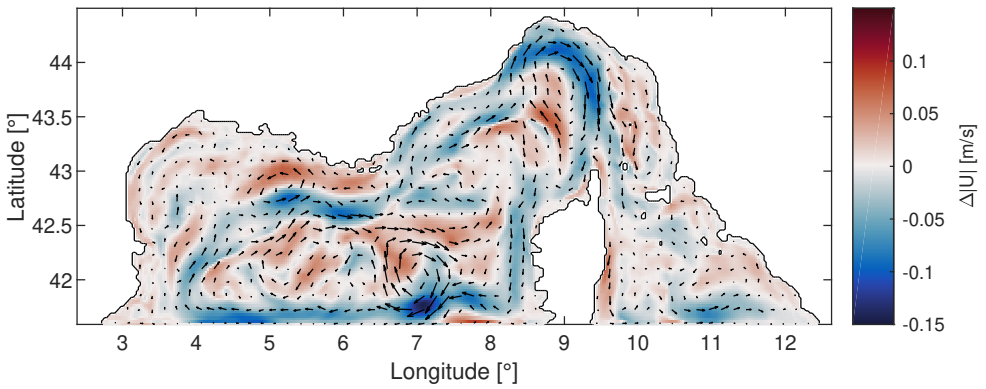


Figure 3: Averaged differences between the upper 50 m flow fields from AN and FR. The arrows indicate the direction of the flow correction. Red is a correction if in favour of the background flow, blue against it.

The difference between the three-month average speed for the FR and the AN experiments is shown in Figure 3 where $\Delta\,|\mathbf{U}| = \sqrt{u_{AN}^2 + v_{AN}^2} - \sqrt{u_{FR}^2 + v_{FR}^2}$. Also shown are vectors of the velocity differences. After data assimilation, an anticyclonic correction is applied on the area North of Corsica, with the flow velocity increasing on the western part and strongly decreasing on the eastern part, downstream the conjunction of ECC and WCC. The NC is mainly modified westward of the TLN HFR region with a cyclonic adjustment close to the French coast and further downstream adjacent to the coast of Spain. Moreover, an anticyclonic increment is detectable westward of Corsica.

To evaluate the skill of the FC and AN runs in reproducing observed currents, with respect to the FR, we computed the root mean squared error $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2}$ and the correlation coefficient $\rho = \frac{\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})}{\hat{\sigma}_y\hat{\sigma}_x}$, where $y_i$ and $x_i$ are observation and model values within a reference temporal window, overbar denotes averages, and $\hat{\sigma}_y$ and $\hat{\sigma}_x$ are the standard deviations of observations and model values, respectively. Furthermore, the centred root mean squared error $cRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[(y_i - \bar{y}) - (x_i - \bar{x})]^2}$ is used within the Taylor Diagrams (Taylor, 2001).

The spatial distribution of $RMSE$ and $\rho$ is shown in Figures 4 and 5. In

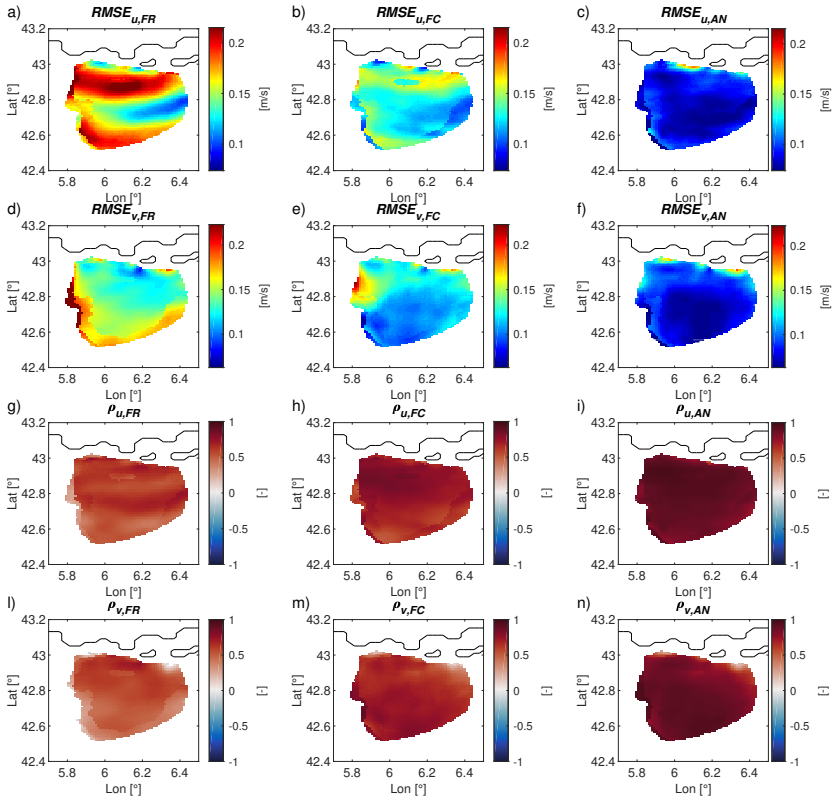


Figure 4: Spatial distribution of $RMSE$ (from a to f) and $\rho$ (from g to n) for the $u$-velocity component and for the $v$-velocity component for the TLN HFR, for free-run (a, d, g and l), forecast (b, e, h and m) and analysis (c, f, i and n).

the following, when we refer to the range of a metric, we indicate the interval between $5^{th}$ and $95^{th}$ percentile considering both velocity components together, unless otherwise stated.

The spatial distribution of $RMSE$ and $\rho$ indicates that the improvement gained through the assimilation procedure is evident for both the $u$- and $v$-velocity components and both the HFR systems. Indeed, the comparison of the $RMSE$ of FR, FC, and AN for the TLN HFR (Figures 4a to 4f) shows a net reduction in the error, especially for $u$. For the most part, the $RMSE$ values range between 0.12 and 0.21 m/s for the FR, between 0.1 and 0.15 m/s for the FC and are below 0.1 m/s for the AN. Concerning the LIG HFR, the trend of

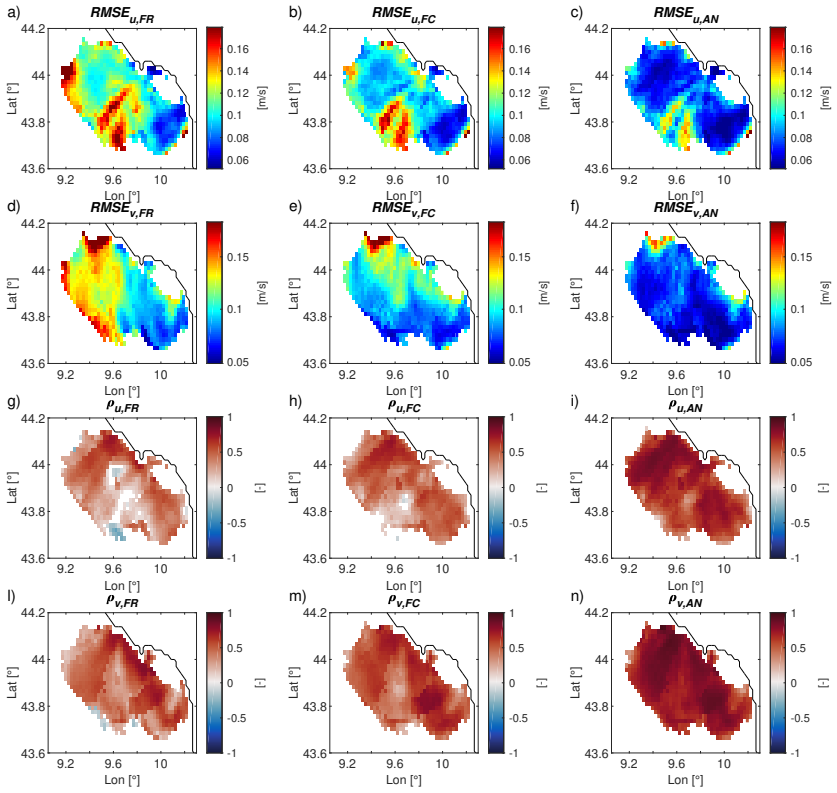


Figure 5: Spatial distribution of $RMSE$ (from a to f) and $\rho$ (from g to n) for the u-velocity component and for the v-velocity component for the LIG HFR, for free-run (a, d, g and l), forecast (b, e, h and m) and analysis (c, f, i and n).

improvement is detectable in Figures 5a to 5f, where the $RMSE$ reduces moving from the FR (0.07-0.17 m/s), to the FC (0.06-0.15 m/s) and AN (0.05-0.12 m/s), for both $u$ and $v$. The western part is the most affected by the assimilation procedure and it is indeed the area corresponding to the anticyclonic increment detectable in Figure 3.

The correlation at the TLN HFR, is in the range 0.32-0.67 for the FR and increases with FC (0.53-0.82) and AN (0.72-0.93, Figure 4g and 4n). Conversely, the correlation at the LIG HFR exhibits lower values than that at TLN HFR, especially for $\rho_{u,FR}$. However, the increase is significant, passing from 0.08-0.68 for the FR, to 0.18-0.70 for the FC, and 0.40-0.86 for the AN.

Globally, considering both the velocity components, the reduction of $RMSE$ for the FC ranges from 14% to 24%, and that of the AN from 33% to 47%. The increase in correlation ranges from 19% to 41% for the FC, and from 35% to 109% for the AN.

To summarize the calculated statistics and the improvement or worsening due to the assimilation procedure, we show in Figure 6 the Taylor Diagrams calculated over the whole dataset, for the $u$- and $v$-velocity components and for the SST, both in the case in which only surface velocities are assimilated, and in the case in which surface velocities and satellite SST are assimilated. The values of standard deviation and $cRMSE$ are normalized with respect to the standard deviation of the observations.

Figure 6a shows the model skill relative to the TLN HFR. The improvement of modelled $u$ (plus sign) and $v$ (circle) velocities for FC and AN with respect to the FR and NM is clear and the assimilation of satellite SST, in addition to HFR surface currents, slightly decreases $cRMSE$ and increases $\rho$. Similarly, the Taylor Diagram relative to the LIG HFR (Figure 6b) shows better model performances after the assimilation procedure, and also in this case, the difference with and without satellite SST assimilation are rather low. When we consider the model skill in reproducing SST, the reduction of error and increase in correlation due to the data assimilation is more clearly detectable for the AN than for the FC. This can be ascribed to the fact that FR was already starting from good model skills. However, the reduction in $cRMSE$ is equal to 12% for FC, and 35% for AN, when SST is assimilated. It is also important to stress that in the case only HFR data are assimilated, the $RMSE$ is increased and the correlation is reduced both for FC and AN. This aspect will be extensively addressed in the Discussion section.

To further validate the results of the DA procedure, model values of velocity are interpolated at hourly averaged drifter locations and then compared to hourly averaged drifter derived velocities, $u_d$ and $v_d$. If the standard deviation of multiple drifter observations in a single hour is greater than 5 times the average value, this is removed from the analysis. The values of $RMSE$ and $\rho$ calculated

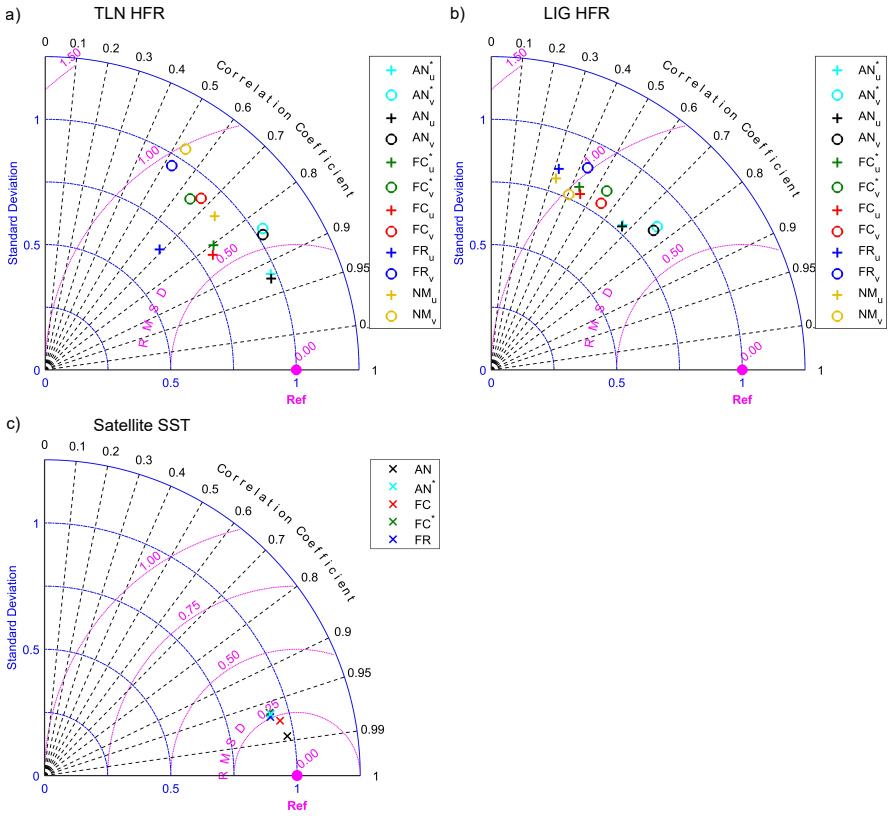


Figure 6: Taylor Diagrams of model performance with respect to assimilated observations in case HFR plus satellite SST data are assimilated, and in case only HFR data are assimilated (legend items with asterisk). Plus signs and circles represent $u$- and $v$-velocity components, respectively, crosses represent SST. a) reports the comparison against TLN HFR data, b) the comparison against LIG HFR data, and c) against satellite SST data. Blue symbols correspond to FR, yellow symbols to NM, red symbols to FC, black symbols to AN, green symbols to FC in case only HFR data are assimilated, and cyan symbols to AN in case only HFR data are assimilated. Model standard deviations and $cRMSE$ are normalized with respect to the standard deviation of observations.

Table 1: Summary of the $RMSE$ and $\rho$ values calculated over all the domain, for the $u$- and $v$-velocity components of the model with respect to surface velocity from surface drifters.

| Run | | | FR | NM | FC | AN |
|---|---|---|---|---|---|---|
| Inside TLN area | $RMSE$ [m/s] | $u$ | 0.216 | 0.271 | 0.184 | 0.142 |
| | | $v$ | 0.15 | 0.152 | 0.133 | 0.101 |
| | $\rho$ [-] | $u$ | 0.74 | 0.433 | 0.812 | 0.88 |
| | | $v$ | 0.575 | 0.379 | 0.548 | 0.806 |
| Inside LIG area | $RMSE$ [m/s] | $u$ | 0.163 | 0.166 | 0.147 | 0.132 |
| | | $v$ | 0.162 | 0.169 | 0.132 | 0.117 |
| | $\rho$ [-] | $u$ | 0.158 | 0.287 | 0.354 | 0.464 |
| | | $v$ | 0.223 | 0.239 | 0.54 | 0.68 |
| Outside HFRs area | $RMSE$ [m/s] | $u$ | 0.152 | 0.154 | 0.155 | 0.153 |
| | | $v$ | 0.158 | 0.156 | 0.158 | 0.158 |
| | $\rho$ [-] | $u$ | 0.61 | 0.594 | 0.591 | 0.612 |
| | | $v$ | 0.594 | 0.581 | 0.588 | 0.58 |
| Whole domain | $RMSE$ [m/s] | $u$ | 0.155 | 0.16 | 0.155 | 0.15 |
| | | $v$ | 0.158 | 0.158 | 0.154 | 0.152 |
| | $\rho$ [-] | $u$ | 0.603 | 0.581 | 0.605 | 0.636 |
| | | $v$ | 0.571 | 0.552 | 0.594 | 0.6 |

within the HFR areas, outside them, and over the whole domain, are reported in Table 1. It is possible to deduce a better performance of both FC and AN with respect to FR and NM specifically inside both the LIG and TLN areas, but for the correlation of the $v$-component in the latter. If we consider the whole domain, the extent of the improvement is small, similarly to the differences in the values of the statistics calculated for the drifters outside the HFRs area, where both improvement and worsening are present (Table 1).

The spatial distribution of the improvement (red cells) and worsening (blue cells) of the FC with respect to the FR is reported in Figure 7a and b for the $RMSE$, and in Figure 7c and d for $\rho$. The analogous comparison between AN and FR is reported in Figure 7e and f for the $RMSE$, and in Figure 7g and h. For the calculation, the domain is divided into cells of 10 by 10 km size. The data contained in each cell are used to calculate the $RMSE$ and $\rho$. Then, the value of $\Delta RMSE$ and $\Delta\rho$ are weighted on the basis of the amount of data

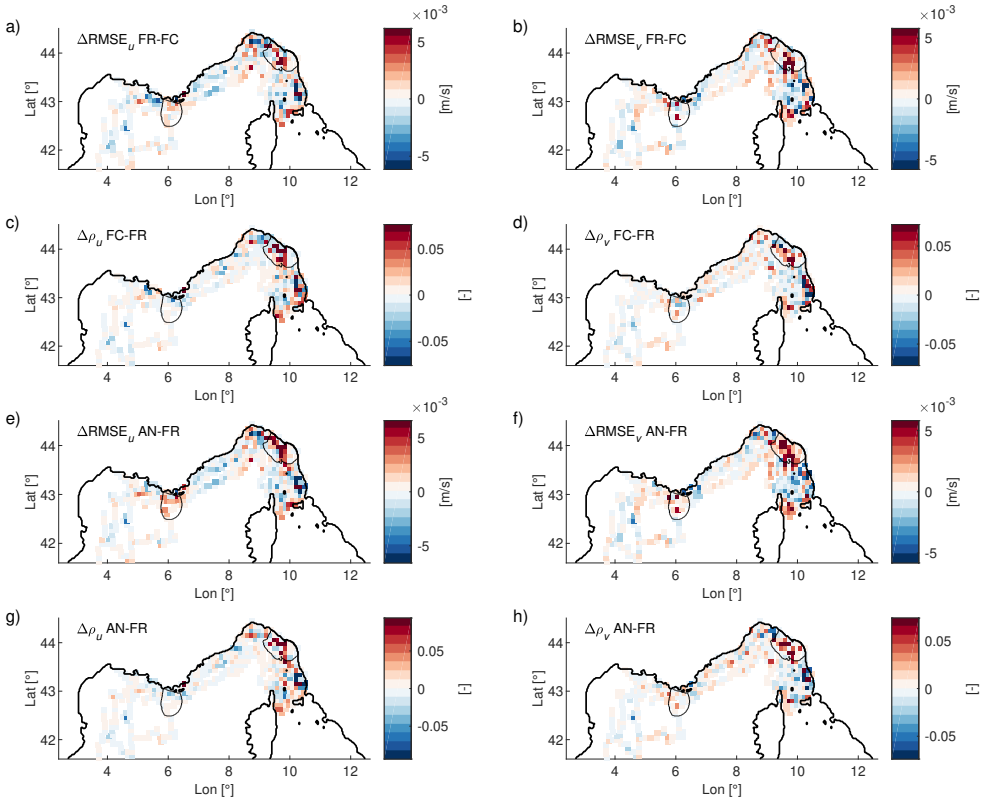


Figure 7: Spatial distribution of the difference of weighted $RMSE$ and $\rho$ for the $u$-velocity component (a and c) and for the $v$-velocity component (b and d) between the FC and FR, and for the $u$-velocity component (e and f) and for the $v$-velocity component (g and h) between the AN and FR. Red colour means an improvement, blue colour a worsening.

contained in each cell. This is to avoid, from a visual point of view, that cells with a few data have the same importance as those with a lot of data.

In Figure 7, we can observe that most of the red cells are located inside the HFR areas, especially in the LIG HFR. Outside those portions of the domain we have regions where the solution is improved and where it is worsen. For both the FC and the AN a worsening is present in the area south of the LIG HFR, but for the Corsica Channel where an improvement is detectable. Soft colours, corresponding to small differences, indicate that there are few data in the cell or the difference in statistics between the compared simulations is small.

A summary of the results is reported in the Taylor Diagrams for the $u$- and $v$-velocities in Figure 8, subdivided by whether or not drifter data are within the HFR areas. Furthermore, we included the comparison of surface velocities from drifters against those from HFR observations, to have an estimate of the agreement between the assimilated observations and those employed to validate the results. Also in this case, the standard deviation and the $cRMSE$ are normalized with respect to the standard deviation of drifter data.

In the following, the increase/decrease of the quality of the model solution by the DA, is computed with respect to the FR only. This is because FR and NM statistics with respect to drifter data are similar within the LIG HFR (Figure 8b) and outside the HFR areas (Figure 8c), whereas the FR performs better then NM within the TLN HFR area (Figure 8a). For the TLN HFR area, the FC improves with respect to the FR for the $u$ component (-9% $cRMSE$ and +10% $\rho$), whereas the $v$ component experience a $cRMSE$ reduction (-12%) but a slight decrease in correlation (-3%). For the AN the reduction in $cRMSE$ (-28% for $u$ and -37% for $v$), and the increase in correlation (+19% for $u$ and +40% for $v$) are more evident. A similar, but more marked behaviour, is also found for the LIG HFR (Figure 8b), but for the low value of the FR statistics as compared to the drifters data. In this case the assimilation procedure decreases the $u$ and $v$ velocity $cRMSE$ (-6% for FC and -14% for AN; -19% for FC and -29% for AN, respectively), and increases the $u$ and $v$ velocity correlations (+120% for FC and +190% for AN; +145% for FC and +208% for AN, respectively). This large increase in correlation is mainly due to the poor performance of the FR in reproducing the drifter derived surface currents in the LIG HFR area. Outside the HFR areas the differences in statistics among the FR, NM, FC and AN are negligible (Figure 8c).

Furthermore, it is worth to be noted that the agreement between drifters data and HFR observations is similar to that between drifter data and AN (green and black symbols in Figure 8a and b), except that the surface velocities measured by HFR have, in general, standard deviations larger than those associated to modelled velocities. This is also true for the most part of the data reported

in Figure 8, whereby surface velocities derived from drifters exhibit the largest standard deviations.

### *3.2. Observation Impact on Alongshore Transport*

The impact of observations in producing variation in the alongshore transport, averaged over each assimilation window, was evaluated for three different transects located close to the HFRs (Tr1 and Tr3) and along a transect in between (Tr2, see Figure 1a and Figure 1a).

The magnitude of the averaged transport is similar along Tr1 and Tr2 and different for the Tr3 (Figures 9a, 9c and 9d) even though the cross-shore distance covered by each section (≈55-60 km for Tr1 and Tr2, and ≈50 km for Tr3) is similar. At Tr1 and Tr2 there is a westward transport of about 0.5 Sv due mainly to the NC, while for Tr3 the transport ranges between -0.1 Sv and 0.1 Sv. In the case of the transport increment $\Delta I$, these are more pronounced for the transects in the vicinity of the two HFRs (Figure 9b, 9d and 9f). Figures 9b, 9d and 9f show the transport increment $\Delta I$ based on tangent linear approximation (blue line) and the non-linear model analysis minus background (black line). The two are very similar, indicating that the tangent linear approximation in Equation 4 is valid.

For all the transects, the three sources of observation have a significant impact on $\Delta I$, with a large contribution of SST even where surface velocities from the TLN and LIG HFRs are assimilated (Figure 10). This aspect will be further addressed in the Discussion section since the statistical results from Figure 6 shows that SST does not affect so much the surface velocities, despite its impact on the average transport. Furthermore, regardless of the magnitude of the transport, both positive and negative increments are present and in general there is consensus between the impact of observations from the different platforms. This means that only in a few cases does one observation type induce an increase in transport while another induces a decrease. Only for Tr3 (Figure 10c) it is possible to observe a general reduction of the transport. This is in agreement with the averaged difference of the upper 50 m flow field between

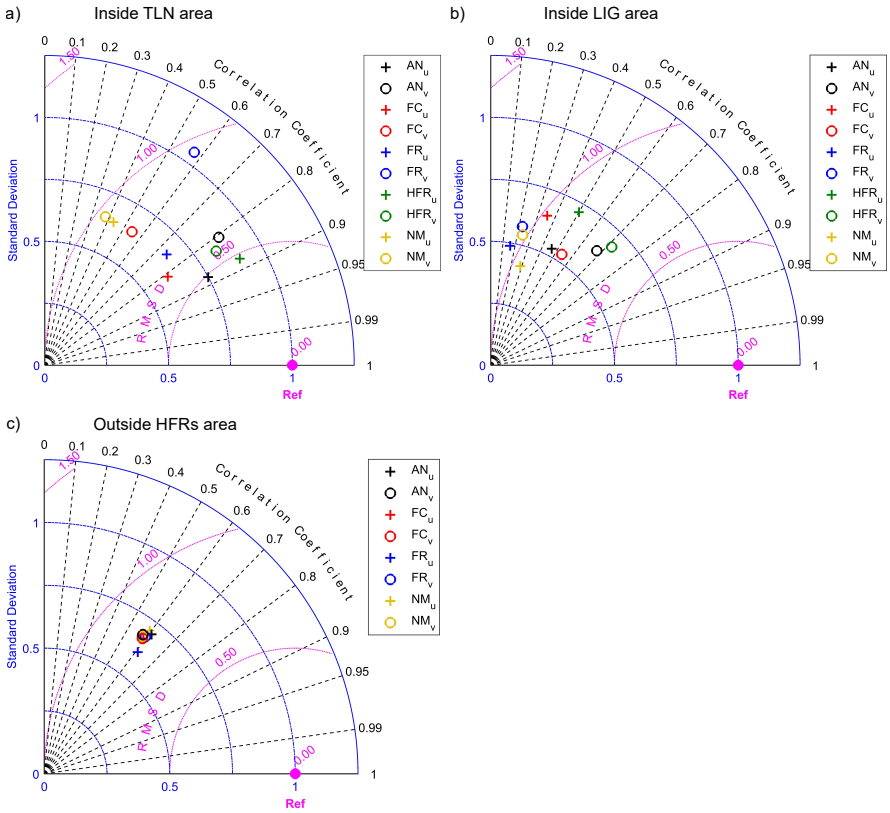


Figure 8: Taylor Diagrams of model and HFR performance with respect to surface velocities from drifters, considering the TLN HFR area a), the LIG HFR area b) and the remaining par of the domain c). Plus signs and circles represent $u$- and $v$-velocity components, respectively. Blue symbols correspond to FR, red symbols to FC, black symbols to AN, and green symbols to HFR observations. Standard deviations and $cRMSE$ are normalized with respect to the standard deviation of drifter velocities.

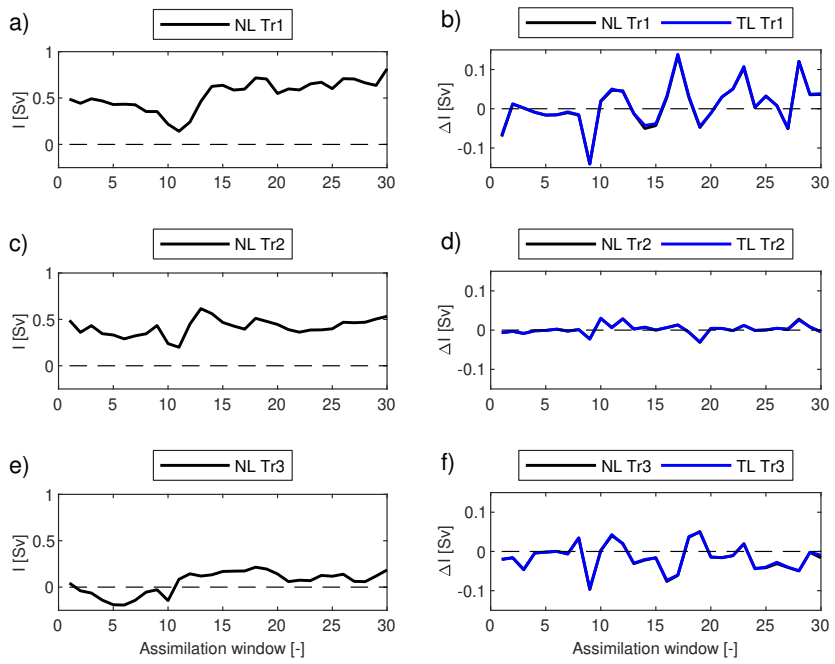


Figure 9: Alongshore volume transport averaged over each assimilation window $I$, for the three transects Tr1 a), Tr2 c) and Tr3 e). Transport increment $\Delta I$ calculated from the non-linear model (black line) and from the tangent Linear model (blue line) for the three transects Tr1 b), Tr2 d) and Tr3 f). Transport is considered positive in the direction of the main current.

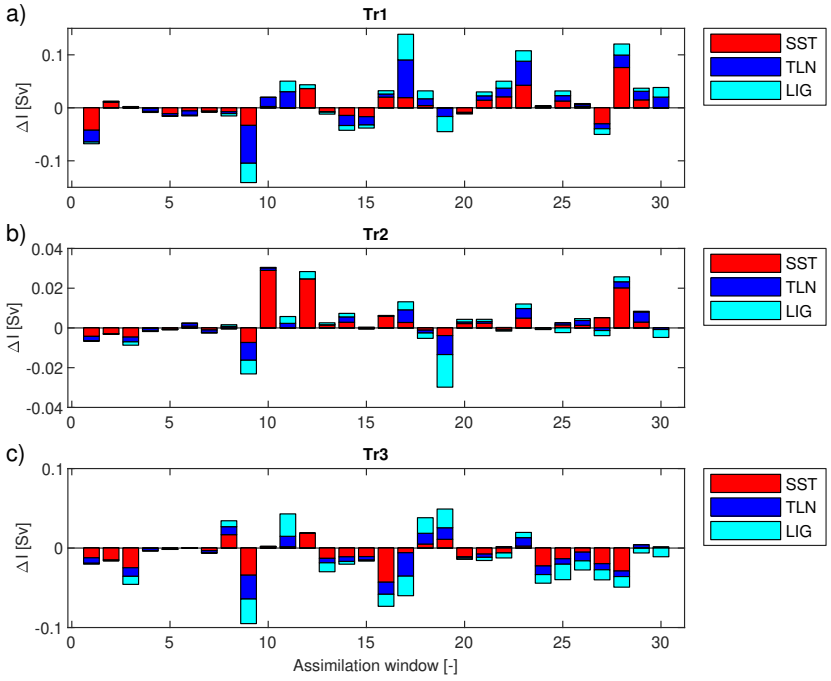


Figure 10: Impact of each source of observation on the transport increment $\Delta I$ for the three transects Tr1 a), Tr2 b) and Tr3 c). The contribute of SST from satellite observations is reported in red, the contribute of surface velocities from TLN HFR in blue and that of surface velocities from LIG HFR is reported in light blue.

the FR and the AN reported in Figure 3.

The impact of each component of the control vector (initial conditions, lateral open boundary conditions, and surface forcing) on the transport increment $\Delta I$ along the three transects, is shown in Figure 11. Tr1 is principally affected by adjustment to the initial condition and atmospheric forcing, whereas the impact of adjustments in the lateral open boundary conditions is negligible (Figure 11a). Farther to the east, along Tr2 and Tr3, the impact of the lateral boundary conditions becomes more significant while there is a relative decrease in the impact of the surface forcing (Figures 11b and 11c). Moreover, even if the transport increment for Tr2 is small (at most 10% of the total transport), in several cases we can observe a competing effect for which the contribution of

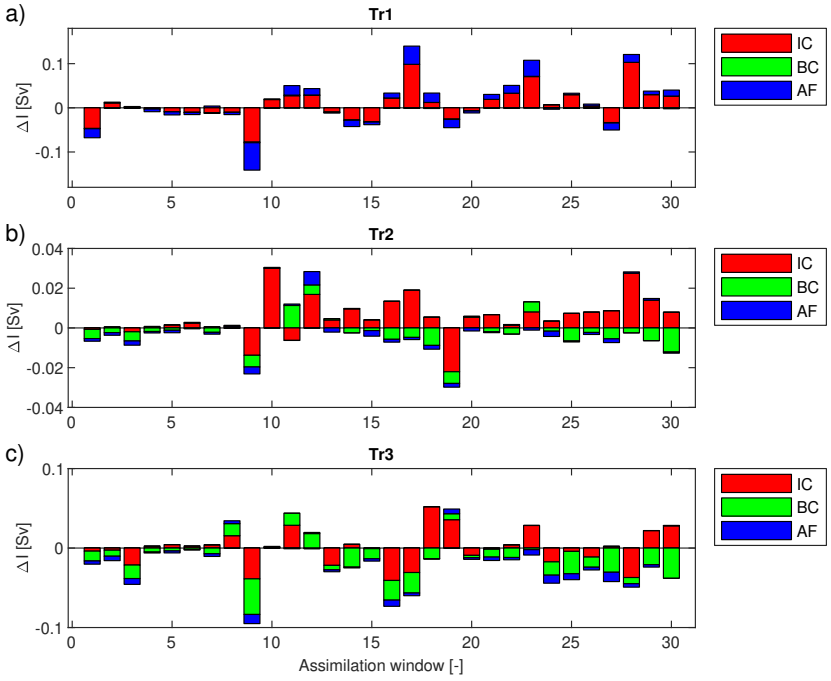


Figure 11: Impact of each element of the control vector on the transport increment $\Delta I$ for the three transects Tr1 a), Tr2 b) and Tr3 c). The contribute of initial conditions (IC) is reported in red, the contribute of boundary conditions (BC) in green, and that of atmospheric forcing (AF) in blue.

initial conditions tends to increase the transport, whereas adjustments to the boundary conditions act in the opposite direction (Figure 11c).

## 4. Discussion

In the present study we assimilated surface velocities from two different HFRs, the TLN and the LIG (Figure 1). In general, the model performs better adjacent to the coast near Toulon than along the coast of Liguria. This might be partly explained by the characteristics of the mean flow field, for which the TLN HFR covers a portion of the NC while the LIG HFR observes an area characterized by a more complex circulation. For the model, it is easier to reproduce a flow pattern with a marked and almost permanent and coherent

large structure, instead a less coherent circulation comprised of eddies and fronts (Figure 1b, 1c, and Figure 1a). The results of the assimilation procedure can be affected by this aspect. Indeed, despite the error in surface velocities for both HFRs is reduced in a similar manner, at least for the $v$-component, the increase in correlation is higher for the LIG HFR than the TLN HFR, for both the FC and AN. This is partially due to the fact that, in the area covered by the LIG HFR, the correlation associated to the FR is quite low (0.31 for $u$-component and 0.43 for the $v$-component) and is more easily increased. Even if the correlation obtained for the FC and AN are lower in this area than in the area covered by the TLN HFR, the improvement is larger both considering the comparison with HFR data (Figure 6) and drifter data (Figure 8).

The error pattern in Figure 5a, b and c may lead to think of some calibration issues for the LIG HFR, although it is present only on the $u$ component of the velocity. Such a pattern remains despite we removed the data with $GDOP \geq 2$, those with no "good" quality flag, and those characterized by large horizontal divergence. This is an aspect that will be further investigated and analysed in the future, for example by assigning a larger error value to the data located in the area or by updating the calibration for the system.

The assimilation of satellite SST, in addition to surface velocities, understandably leads to an improvement in the modelled SST. When only HFR data were assimilated, the overall $RMSE$ was higher, and the correlation slightly lower, for both FC (+7% and -5%) and AN (+4% and -3%) with respect to the FR, compared to the case when SST are assimilated. This effect was also observed by Zhang et al. (2010), but for subsurface dynamics which they attributed to deficiencies in the background error covariance, over-correction of the control vector or limitation of the non-linear model. Here, we suggest a simple measure to reduce the worsening of the modelled SST pattern, if satellite data were not assimilated. The background error covariance matrix for the $T$ and surface heat flux, was determined by means of a standard deviation calculated after the seasonal trend and the diurnal signal were removed from the time series. The choice was driven by the need to reduce the entity of the

background error for the two quantities, which principally affect the SST field. After the filtering procedure, we obtained median values of standard deviation for the upper layer equal to 0.74 °C and 59 W/m$^2$ for $T$ and the surface heat flux, respectively. Before this, the median values were equal to 2.55 °C and 111 W/m$^2$ and lead to $RMSE$ increase up to 40% for both FC and AN, when only surface velocities from HFRs were assimilated. In addition, this choice left almost unchanged the improvement in the FC and AN for the modelled SST, in case satellite observations were assimilated.

The comparison of modelled results against non assimilated observations allows us to further assess the convenience of the data assimilation procedure. Furthermore, it is possible to identify a sort of limit concerning the skill of the assimilated model to reproduce independent observations, that is the degree of agreement between the independent observation themselves and those assimilated. In Figure 8a and b we can see that the AN is quite close to the HFR statistical performances and we cannot expect to get closer than that to drifter data.

Outside the area where surface velocities are assimilated (Figure 8a) the assimilation procedure is not able to clearly improve the modelled circulation, and there are several regions where the FR is more in agreement with drifters data than FC and AN (Figure 7). However, it is important to take into account that the increase in error could be due to several factors: deficiencies in the background error covariance matrix and in the observation error covariance matrix, violation of the tangent linear assumption, the low quality of the assimilated observations or the mismatch between the assimilated observations and those employed to validate the results (drifters). Indeed, while the agreement between the two dataset is high for the TLN HFR (Figure 8a), with $\rho$ larger than 0.8 and normalized $RMSE$ around 0.5, this is not equally true for the LIG HFR (Figure 8b), where the correlations are equal to 0.5 ($u$) and 0.7 ($v$) and the normalized $RMSE$ is 0.8 for the $u$-component and 0.8 for the $v$-component.

The contribution of observations from different platforms to the variation of the upper surface alongshore transport was analysed at three transects. One

particular aspect is the significant effect of SST on the alongshore transport (Figure 10), while the differences in statistics between the case with and without SST assimilation are negligible (Figure 6). To shed light on this aspect, we determined the observation impact on the transport increment at Tr1 in case only surface velocities were assimilated (Figure 12b).

Transport increments $\Delta I$ on Tr1 are very similar regardless of whether satellite SST is assimilated (Figure 12a). This information also appears to contradict the significant effect that assimilated satellite SST has on $\Delta I$ (Figure 10a). To explain this, we considered the $28^{\text{th}}$ assimilation window, from 2020-11-21 at 00:00 to 2020-11-23 at 23:59 (circled in red in Figure 12b), for which the $\Delta I$ is almost equal (0.125 Sv with SST and 0.124 Sv without SST), but the contribution of SST is larger than that of surface velocities in the case satellite data are assimilated (Figure 10a). During this time window, we determined the vertical distribution of the averaged difference $\delta u = u_a - u_b$, of the analysis ($a$ subscript) and background ($b$ subscript) $u$-velocity component on the Tr1 transect both in case satellite SST is assimilated (Figure 12c) or not (Figure 12e). Similarly, we determined the spatial distribution of averaged value of $\delta u$ over the upper 50 m all over the domain (Figures 12d with SST and 12f without SST). It is important to keep in mind that, as part of the present discussion, we refer to "background" as the non-linear model initialized from the analysis at the end of the previous assimilation window, and as "analysis", the first outer loop of the 4D-Var DA procedure (see Section 2). Figures 12c and 12e show that the effect of assimilated SST is to move the increment more toward the coast and change the distribution of the increment itself. It is then clear that SST has a significant impact on the distribution of the velocity difference $\delta u$, while maintaining almost unchanged the transport increment. Furthermore, the spatial distribution of the upper 50 m $\delta u$ in case satellite SST is assimilated (Figure 12d) or not (Figure 12f), shows that the ingestion of SST, covering the entire domain, affects the whole upper surface dynamic, whereas the HFR observations mainly influence the areas where they are located, and their surroundings. The volume transport is calculated by means of the local $u$-velocity component, which de-

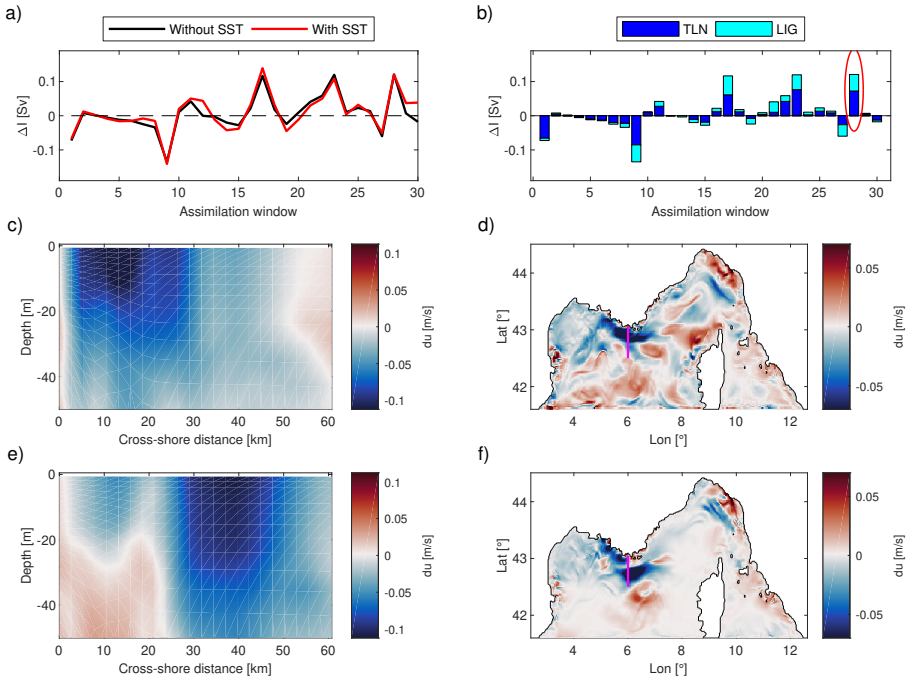


Figure 12: a) Transport increment $\Delta I$ at the Tr1 transect in case satellite SST is assimilated (red line) or not (black line). b) Contribute of observation typology to transport increment $\Delta I$ in case only surface velocities from TLN HFR (blue) and LIG HFR (light blue) are assimilated. Red ellipse indicates the $28^{\text{th}}$ assimilation window from 2020-11-21 00:00 to 2020-11-23 23:59. c) Averaged value of the difference $\delta u = u_a - u_b$ on the Tr1 transect, up to 50 m depth, for the assimilation window encircled in b) in case satellite SST is assimilated; e) in case SST is not assimilated. d) Spatial distribution of the depth-averaged (50 m) value of the difference $\delta u = u_a - u_b$ for the assimilation window encircled in b) in case satellite SST is assimilated; f) in case it is not assimilated.

pends on the the values of $u$ on the remaining part of the domain, which, in turn, is affected by the assimilated SST. This explains the significant effect of the latter on $\Delta I$.

Another noteworthy aspect is the comparison of $\Delta I$ at Tr1 and Tr2. The former is indeed larger than the latter, while the transport $I$ is similar. In the area of Tr1 we have both surface velocities and SST, whereas in the Tr2 area only SST (Figure 1a). Based on the significant impact of SST at Tr1 with $\Delta I$ ranging between -0.12 and 0.12 Sv, we should expect at Tr2 a larger $\Delta I$ than observed. This behaviour can be explained if we consider that SST alone does not affect the alongshore transport, but intervenes once it is modified by assimilated surface velocities. Through the combined effect of the adjoint model and the background error covariances the contribution of SST is enabled once a substantial modification to the flow field is induced.

Lastly, we have to keep in mind that 4D-Var DA does not act linearly on the increment of the control vector $\delta\mathbf{z}$. Instead, it operates in a non-linear way by means of the Kalman gain matrix $\mathbf{K}$, which mutually weights observation and background error covariances, and propagates information backward in time through the adjoint, leading to results that can be difficult to interpret dynamically.

In addition, the structure of the background error covariance matrix is fundamental in determining the increment $\delta\mathbf{z}$, which is a linear combination of the columns of $\mathbf{D}$. To rely on a consolidated approach, we assumed $\mathbf{D}$ to be proportional to the climatological variability of the nonlinear model, as employed in previous studies (Moore et al., 2011a; Iermano et al., 2016; Phillipson and Toumi, 2017; Janeković et al., 2020). However, we are aware that different approaches to determine $\mathbf{D}$, for example by standard deviations associated to the last 2-3 weeks prior to the assimilation window, or through the so called NMC approach (Parrish and Derber, 1992), may lead to different results.

Figure 11 reports the impact of each component of the control vector on the transport increment. The negligible contribution of boundary conditions on $\Delta I$ at Tr1 and its increasing contribution moving toward Tr2 and Tr3, can

be explained by the fact that the main flow pattern is cyclonic, and brings information from the boundaries to the areas where the transects are located, by advection and wave propagation.

Moreover, it is possible to observe the persistence of the impact of initial conditions during the whole set of assimilation windows. On the one hand we may think that after subsequent corrections the system has lost its memory of the freerun and the impact of the initial conditions should reduce. On the other hand, being the assimilation window of three days, the initial condition still plays a crucial role in determine the dynamic of the system, and modifications to such a dynamic have necessarily to come from them. Lastly, the three months period should be too short to highlight a reduction in the importance of the initial condition in modifying the surface transport.

The way the assimilation procedure corrects the background can be analysed from the point of view of the geostrophic and ageostrophic contributions to the surface velocity. To this aim, we determined the root mean squared difference ($RMSD$) between the background (i.e. FC) and analysis runs for the velocity magnitude $|\mathbf{U}|$. The statistic is calculated over the whole set of assimilation windows. The geostrophic contribution to the surface velocity is calculate as $u_g = -\frac{g}{f}\frac{\partial \eta}{\partial y}$ and $v_g = \frac{g}{f}\frac{\partial \eta}{\partial x}$, with $\eta$ the modelled sea surface, $g$ the gravitational acceleration and $f$ the Coriolis parameter. The ageostrophic contribution is simply determined as $u_a = u - u_g$ and $v_a = v - v_g$. Figure 13 show the spatial distribution of the $RMSD$ between the background and analysis runs. It is possible to observe that the assimilation procedure modifies the ageostrophic surface velocity (Figure 13b) more than the geostrophic part (Figure 13a). This is indeed in agreement with what reported by Hernandez-Lasheras et al. (2021) and Yu et al. (2012), for which it is the altimetry that better constraints the geostrophic circulation. Moreover, the modifications tend to be located all over the domain and more concentrated in correspondence of the HFR areas, both for the ageostrophic and geostrophic parts, albeit to a lesser extent for the latter (Figure 13a and b).

Another noteworthy aspect is the presence of potential discontinuities be-

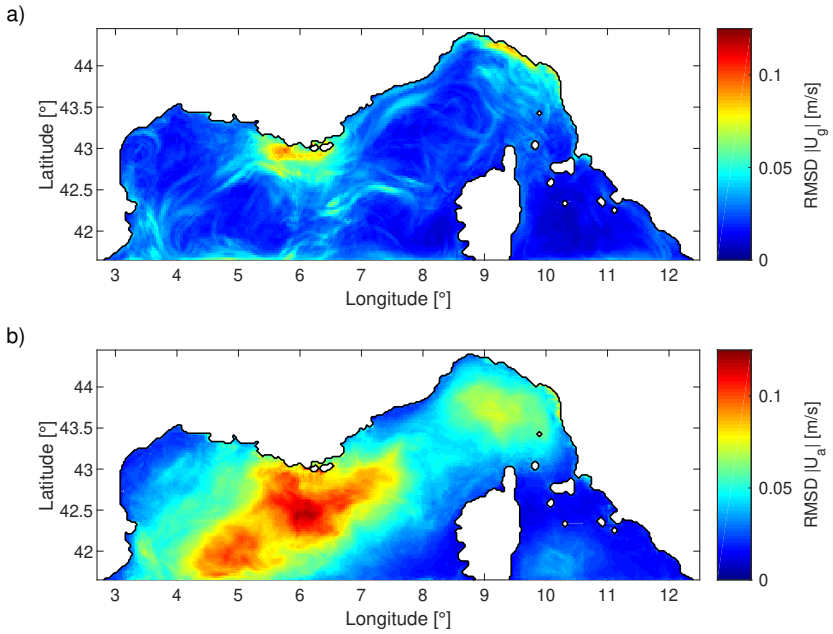


Figure 13: Root mean squared difference between the background and analysis runs for the geostrophic $|\mathbf{U}_g|$ a), and ageostrophic $|\mathbf{U}_a|$ b), velocity magnitudes.

tween subsequent assimilation cycles and the presence of initialization shocks (Balmaseda et al., 2009; Moore et al., 2019). This aspect is widely addressed in the realm of atmosphere, by means for example, of digital filter initialisation (DFI), (Peckham et al., 2016). In the ocean it is generally addressed when ocean fields are used to drive biogeochemical models, since they are negatively impacted by spurious vertical velocities caused by the presence of initialization shocks (Waters et al., 2017). Nevertheless, with reference to the objectives of the preset work, we are more interested in the final portion of the assimilation window, since it provides the initial condition for the subsequent forecast. In fact, this is less affected by the shocks that are dissipated throughout the simulation.

Considering the employed modelling setup, a potential limitation of the current approach is the lack of tidal forcing, that can have an effect on the surface velocities. Nevertheless, we argue that it does not substantially change the main findings and results of the present work, and in fact it is not even included in those of Hernandez-Lasheras et al. (2021) and Vandenbulcke et al. (2017). This aspect will be tackled in the next version of the data assimilation system.

The assimilation procedure also affects the salinity field even if no salinity observations are assimilated. This is especially true at the river mouths which are characterized by larger standard deviations and, as a consequence, larger background error values with respect to the remaining part of the domain. The salinity increments (not shown) are related the initial condition, the boundary conditions and the atmospheric forcing, but they are not directly linked to the river inputs. This might be a shortcoming of the present approach and deserves to be investigated in future developments.

## 5. Conclusion

In the present work we analyse the effect and the impact of the assimilation of surface velocities from two HFR systems and the SST from satellite on a ROMS model implemented on the North-Western Mediterranean Sea. For a

three months period we assimilated data over 3-day windows by means of the ROMS 4D-Var, correcting initial conditions, boundary conditions and atmospheric forcing. A free run without data assimilation is employed as reference solution to quantify the improvement gained by DA. Every 3 days a forecast run is performed with initial condition coming from the final time of the previous analysis run.

The comparison between modelled and surface velocities observed by HFRs shows a net improvement in the description of surface circulations, both for the AN (analysis) and the FC (forecast) with respect to the FR (freerun) and NM (NEMO) solutions. The reduction of the root-mean-squared error $RMSE$ and the increase in correlation $\rho$ is higher in the AN than in the FC.

The assimilation of SST improves both forecast and analysis solution in reproducing the surface dynamics of both SST and, to a lesser extent, surface velocity. In addition, it prevents the degradation of the SST pattern in the case where only surface velocities are assimilated.

The validation of the assimilation procedure against independent observations of surface velocities derived from Lagrangian drifters, shows that the solution is improved within the HFR areas with respect to the FR and NM. Furthermore, the AN shows error and correlation values close to that of the HFR when we compare the two sources of observations of surface velocities. This highlights the existence of an upper limit for the DA skills in improving surface circulation, and it is strongly dependent on the quality and characteristics of the assimilated observation. In the remaining part of the domain, the surface circulation is modified giving both improvements and deteriorations that balance each other, leaving the statistics for FR, FC and AN almost unchanged.

The impact of assimilated observations on the alongshore mass transport on the upper 50 m is evaluated on three transects, two (Tr1 and Tr3) located at the HFR systems and one in between (Tr2), along the flow of the Northern Current. Variations on the mean transport induced by the assimilation procedure are more marked in vicinity of the two HFRs (roughly an order of magnitude larger for the Tr1 and Tr3 than for the Tr2), resulting in variations

that alternate between an increase and reduction of the mean transport, especially for the westernmost transect (Tr1). The impact of the SST is significant and acts to further correct the velocity distribution of the increment induced by the HFRs observations, while it does not change the overall value of the transport increment. Initial conditions play a major role in affecting the alongshore transport at all the analysed sites. Boundary conditions are negligible for the westernmost transect, increasing importance moving eastward, whereas the atmospheric forcing play a role comparable to initial conditions only at the Tr1 transect.

The comparison of the ageostrophic and geostrophic components of the surface velocity, for the background and analysis, shows that the assimilation of HFR and SST data is more effective in modifying the ageostrophic currents rather than the geostrophic.

The present work shows that the improvement for the FC solution is promising from the point of view of implementing an operational system and lays the foundation for the possibility to assimilate further observations such as in situ temperature and salinity profiles and sea surface height. Furthermore, in the future, the useful tool quantifying the observation impact (Moore et al., 2011b) can also be applied to a metric describing the mismatch between modelled and drifter-derived velocities and locations to detect which HFR observations are responsible of the degrading of the solution. This will shed light on the complex issue concerning the choice of the appropriate observations to be assimilated.

## Conflict of interest

The authors declare that they have no conflict of interest.


## Funding information

This research was partly funded by the EU SICOMAR Plus project (2018-2021).

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## References


Balmaseda, M.A., Alves, O.J., Arribas, A., Awaji, T., Behringer, D.W., Ferry, N., Fujii, Y., Lee, T., Rienecker, M., Rosati, T., et al., 2009. Ocean initialization for seasonal forecasts. Oceanography 22, 154–159.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2003. Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC.

Barrick, D.E., Lipa, B.J., 1985. Mapping surface currents. Sea Technol 26, 43–48.

Bellomo, L., Griffa, A., Cosoli, S., Falco, P., Gerin, R., Iermano, I., Kalampokis, A., Kokkini, Z., Lana, A., Magaldi, M., et al., 2015. Toward an integrated hf radar network in the mediterranean sea to improve search and rescue and oil spill response: the tosca project experience. Journal of Operational Oceanography 8, 95–107.

Berta, M., Poulain, P.M., Sciascia, R., Griffa, A., Magaldi, M., 2021. Carthe drifters deployment within the ddr20 - "drifter demonstration and research 2020" experiment in the nw mediterranean sea .

Bourg, N., Molcard, A., 2021. Northern boundary current variability and mesoscale dynamics: a long-term hf radar monitoring in the north-western mediterranean sea. Ocean Dynamics 71, 851–870.

Breivik, O., Saetra, O., 2001. Real time assimilation of hf radar currents into a coastal ocean model. Journal of Marine Systems 28, 161–182.

Chapman, D.C., 1985. Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model. Journal of Physical oceanography 15, 1060–1075.

Clementi, E., Aydogdu, A., Goglio, A., Pistoia, J., Escudier, R., Drudi, M., Grandi, A., Mariani, A., Lyubartsev, V., Lecci, R., et al., 2021. Mediterranean sea physical analysis and forecast (cmems med-currents, eas6 system)(version 1)[data set]. Copernicus Monitoring Environment Marine Service (CMEMS). doi 10.

Courtier, P., 1997. Dual formulation of four-dimensional variational assimilation. Quarterly Journal of the Royal Meteorological Society 123, 2449–2461.

Courtier, P., Thépaut, J.N., Hollingsworth, A., 1994. A strategy for operational implementation of 4d-var, using an incremental approach. Quarterly Journal of the Royal Meteorological Society 120, 1367–1387.

Couvelard, X., Messager, C., Penven, P., Smet, S., Lattes, P., 2021. Benefits of radar-derived surface current assimilation for south of africa ocean circulation. Geoscience Letters 8, 1–11.

Cucco, A., Daniel, P., 2016. Numerical modeling of oil pollution in the western mediterranean sea. Oil Pollution in the Mediterranean Sea: Part I , 255–274.

Dumas, D., Gramoullé, A., Guérin, C.A., Molcard, A., Ourmières, Y., Zakardjian, B., 2020. Multistatic estimation of high-frequency radar surface currents in the region of toulon. Ocean Dynamics 70, 1485–1503.

Fairall, C.W., Bradley, E.F., Rogers, D.P., Edson, J.B., Young, G.S., 1996. Bulk parameterization of air-sea fluxes for tropical ocean-global atmosphere coupled-ocean atmosphere response experiment. Journal of Geophysical Research: Oceans 101, 3747–3764.

Fang, Y.C., Weingartner, T.J., Potter, R.A., Winsor, P.R., Statscewich, H., 2015. Quality assessment of hf radar–derived surface currents using optimal interpolation. Journal of Atmospheric and Oceanic Technology 32, 282–296.

Flather, R., 1976. A tidal model of the north-west european continental shelf. Memoires de la Societe Royale de Sciences de Liege 6, 141–164.

Fossi, M.C., Romeo, T., Baini, M., Panti, C., Marsili, L., Campani, T., Canese, S., Galgani, F., Druon, J.N., Airoldi, S., et al., 2017. Plastic debris occurrence, convergence areas and fin whales feeding ground in the mediterranean marine protected area pelagos sanctuary: a modeling approach. Frontiers in marine science , 167.

Guérin, C.A., Dumas, D., Gramoulle, A., Quentin, C., Saillard, M., Molcard, A., 2019. The multistatic oceanographic hf radar network in toulon, in: 2019 International Radar Conference (RADAR), IEEE. pp. 1–5.

Gurgel, K.W., Antonischki, G., Essen, H.H., Schlick, T., 1999. Wellen radar (wera): A new ground-wave hf radar for ocean remote sensing. Coastal engineering 37, 219–234.

Gürol, S., Weaver, A.T., Moore, A.M., Piacentini, A., Arango, H.G., Gratton, S., 2014. B-preconditioned minimization algorithms for variational data assimilation with the dual formulation. Quarterly Journal of the Royal Meteorological Society 140, 539–556.

Gurvan, M., Bourdallé-Badie, R., Chanut, J., Clementi, E., Coward, A., Ethé, C., Iovino, D., Lea, D., Lévy, C., Lovato, T., Martin, N., Masson, S., Mocavero, S., Rousset, C., Storkey, D., Müeller, S., Nurser, G., Bell, M., Samson, G., Mathiot, P., Mele, F., Moulin, A., 2022. Nemo ocean engine. URL: `https://doi.org/10.5281/zenodo.6334656`, doi:`10.5281/zenodo.6334656`.

Haidvogel, D.B., Arango, H.G., Hedstrom, K., Beckmann, A., Malanotte-Rizzoli, P., Shchepetkin, A.F., 2000. Model evaluation experiments in the north atlantic basin: simulations in nonlinear terrain-following coordinates. Dynamics of atmospheres and oceans 32, 239–281.

Haney, R.L., 1991. On the pressure gradient force over steep topography in sigma coordinate ocean models. Journal of physical Oceanography 21, 610–619.

Hernandez-Lasheras, J., Mourre, B., Orfila, A., Santana, A., Reyes, E., Tintoré, J., 2021. Evaluating high-frequency radar data assimilation impact in coastal ocean operational modelling. Ocean Science 17, 1157–1175.

Houpert, L., Testor, P., De Madron, X.D., Somot, S., D'ortenzio, F., Estournel, C., Lavigne, H., 2015. Seasonal cycle of the mixed layer, the seasonal thermocline and the upper-ocean heat storage rate in the mediterranean sea derived from observations. Progress in Oceanography 132, 333–352.

Iermano, I., Moore, A., Zambianchi, E., 2016. Impacts of a 4-dimensional variational data assimilation in a coastal ocean model of southern tyrrhenian sea. Journal of Marine Systems 154, 157–171.

Janeković, I., Mihanović, H., Vilibić, I., Grčić, B., Ivatek-Šahdan, S., Tudor, M., Djakovac, T., 2020. Using multi-platform 4d-var data assimilation to improve modeling of adriatic sea dynamics. Ocean Modelling 146, 101538.

Janeković, I., Powell, B., Matthews, D., McManus, M., Sevadjian, J., 2013. 4d-var data assimilation in a nested, coastal ocean model: A hawaiian case study. Journal of Geophysical Research: Oceans 118, 5022–5035.

Langland, R.H., Baker, N.L., 2004. Estimation of observation impact using the nrl atmospheric variational data assimilation adjoint system. Tellus A: Dynamic Meteorology and Oceanography 56, 189–201.

Levin, J., Arango, H.G., Laughlin, B., Hunter, E., Wilkin, J., Moore, A.M., 2020. Observation impacts on the mid-atlantic bight front and cross-shelf transport in 4d-var ocean state estimates: Part i—multiplatform analysis. Ocean Modelling 156, 101721.

Levin, J., Arango, H.G., Laughlin, B., Wilkin, J., Moore, A.M., 2021. The impact of remote sensing observations on cross-shelf transport estimates from 4d-var analyses of the mid-atlantic bight. Advances in space research 68, 553–570.

Lorente, P., Aguiar, E., Bendoni, M., Berta, M., Brandini, C., Cáceres-Euse, A., Capodici, F., Cianelli, D., Ciraolo, G., Corgnati, L., et al., 2022. Coastal high-frequency radars in the mediterranean–part 1: Status of operations and a framework for future development. Ocean Science 18, 761–795.

Marmain, J., Molcard, A., Forget, P., Barth, A., Ourmieres, Y., 2014. Assimilation of hf radar surface currents to optimize forcing in the northwestern mediterranean sea. Nonlinear Processes in Geophysics 21, 659–675.

Millot, C., 1999. Circulation in the western mediterranean sea. Journal of Marine Systems 20, 423–442.

Molcard, A., Gramoullé, A., Mazoyer, C., Bourg, N., Ourmières, Y., 2021. Dynamics and transport from the boundary northern current toward the toulon bay: multi-platform observations and downscaling modelling approaches. Ocean Dynamics 71, 993–1009.

Moore, A.M., Arango, H.G., Broquet, G., Edwards, C., Veneziani, M., Powell, B., Foley, D., Doyle, J.D., Costa, D., Robinson, P., 2011a. The regional ocean modeling system (roms) 4-dimensional variational data assimilation systems: part ii–performance and application to the california current system. Progress in Oceanography 91, 50–73.

Moore, A.M., Arango, H.G., Broquet, G., Edwards, C., Veneziani, M., Powell, B., Foley, D., Doyle, J.D., Costa, D., Robinson, P., 2011b. The regional ocean modeling system (roms) 4-dimensional variational data assimilation systems: Part iii–observation impact and observation sensitivity in the california current system. Progress in Oceanography 91, 74–94.

Moore, A.M., Arango, H.G., Broquet, G., Powell, B.S., Weaver, A.T., Zavala-Garay, J., 2011c. The regional ocean modeling system (roms) 4-dimensional variational data assimilation systems: Part i–system overview and formulation. Progress in Oceanography 91, 34–49.

Moore, A.M., Edwards, C.A., Fiechter, J., Drake, P., Neveu, E., Arango, H.G., Gürol, S., Weaver, A.T., 2013. A 4d-var analysis system for the california current: A prototype for an operational regional ocean data assimilation system, in: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II). Springer, pp. 345–366.

Moore, A.M., Jacox, M.G., Crawford, W.J., Laughlin, B., Edwards, C.A., Fiechter, J., 2017. The impact of the ocean observing system on estimates of the california current circulation spanning three decades. Progress in Oceanography 156, 41–60.

Moore, A.M., Martin, M.J., Akella, S., Arango, H.G., Balmaseda, M., Bertino, L., Ciavatta, S., Cornuelle, B., Cummings, J., Frolov, S., et al., 2019. Synthesis of ocean observations using data assimilation for operational, real-time and reanalysis systems: A more complete picture of the state of the ocean. Frontiers in Marine Science 6, 90.

Nardelli, B.B., Tronconi, C., Pisano, A., 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. Remote Sensing of Environment 129, 1–16.

Novelli, G., Guigand, C.M., Cousin, C., Ryan, E.H., Laxague, N.J., Dai, H., Haus, B.K., Özgökmen, T.M., 2017. A biodegradable surface drifter for ocean sampling on a massive scale. Journal of Atmospheric and Oceanic Technology 34, 2509–2532.

Paduan, J.D., Shulman, I., 2004. Hf radar data assimilation in the monterey bay area. Journal of Geophysical Research: Oceans 109.

Parrish, D.F., Derber, J.C., 1992. The national meteorological center's spectral statistical-interpolation analysis system. Monthly Weather Review 120, 1747–1763.

Partridge, D., Friedrich, T., Powell, B.S., 2019. Reanalysis of the pacioos hawaiian island ocean forecast system, an implementation of the regional ocean modeling system v3. 6. Geoscientific Model Development 12, 195–213.

Pawlowicz, R., 2020. M_map: A mapping package for matlab, version 1.4 m. Computer software, UBC EOAS, available at: https://www. eoas. ubc. ca/˜ rich/map. html, last access 29.

Peckham, S.E., Smirnova, T.G., Benjamin, S.G., Brown, J.M., Kenyon, J.S., 2016. Implementation of a digital filter initialization in the wrf model and its application in the rapid refresh. Monthly Weather Review 144, 99–106.

Phillipson, L., Toumi, R., 2017. Impact of data assimilation on ocean current forecasts in the angola basin. Ocean Modelling 114, 45–58.

Quattrocchi, G., Sinerchia, M., Colloca, F., Fiorentino, F., Garofalo, G., Cucco, A., 2019. Hydrodynamic controls on connectivity of the high commercial value shrimp parapenaeus longirostris (lucas, 1846) in the mediterranean sea. Scientific reports 9, 1–14.

Reyes, E., Aguiar, E., Bendoni, M., Berta, M., Brandini, C., Cáceres-Euse, A., Capodici, F., Cardin, V., Cianelli, D., Ciraolo, G., et al., 2022. Coastal high-frequency radars in the mediterranean–part 2: Applications in support of science priorities and societal needs. Ocean Science 18, 797–837.

Reyes, E., Hernández-Carrasco, I., Révelard, A., Mourre, B., Rotllán, P., Comerma, E., Bakhsh, T., Rubio, A., Mader, J., Ferrer, L., et al., 2020. Ibisar service for real-time data ranking in the ibi area for emergency responders and sar operators. Copernicus Marine Service Ocean State Report , 92–99.

Rossi, V., Ser-Giacomi, E., López, C., Hernández-García, E., 2014. Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves. Geophysical Research Letters 41, 2883–2891.

Ser-Giacomi, E., Rossi, V., López, C., Hernandez-Garcia, E., 2015. Flow networks: A characterization of geophysical fluid transport. Chaos: An Interdisciplinary Journal of Nonlinear Science 25, 036404.

Shchepetkin, A.F., McWilliams, J.C., 2003. A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate. Journal of Geophysical Research: Oceans 108.

Shchepetkin, A.F., McWilliams, J.C., 2005. The regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean modelling 9, 347–404.

Sikirić, M.D., Janeković, I., Kuzmić, M., 2009. A new approach to bathymetry smoothing in sigma-coordinate ocean models. Ocean Modelling 29, 128–136.

Soto-Navarro, J., Jordá, G., Deudero, S., Alomar, C., Amores, Á., Compa, M., 2020. 3d hotspots of marine litter in the mediterranean: A modeling study. Marine Pollution Bulletin 155, 111159.

Sperrevik, A.K., Christensen, K.H., Röhrs, J., 2015. Constraining energetic slope currents through assimilation of high-frequency radar observations. Ocean Science 11, 237–249.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres 106, 7183–7192.

Trémolet, Y., 2008. Computation of observation sensitivity and observation impact in incremental variational data assimilation. Tellus A: Dynamic Meteorology and Oceanography 60, 964–978.

Van Sebille, E., Aliani, S., Law, K.L., Maximenko, N., Alsina, J.M., Bagaev, A., Bergmann, M., Chapron, B., Chubarenko, I., Cózar, A., et al., 2020. The physical oceanography of the transport of floating marine debris. Environmental Research Letters 15, 023003.

Vandenbulcke, L., Beckers, J.M., Barth, A., 2017. Correction of inertial oscillations by assimilation of hf radar data in a model of the ligurian sea. Ocean Dynamics 67, 117–135.

Warner, J.C., Sherwood, C.R., Arango, H.G., Signell, R.P., 2005. Performance of four turbulence closure models implemented using a generic length scale method. Ocean Modelling 8, 81–113.

Waters, J., Bell, M., Martin, M., Lea, D., 2017. Reducing ocean model imbalances in the equatorial region caused by data assimilation. Quarterly Journal of the Royal Meteorological Society 143, 195–208.

Weaver, A., Courtier, P., 2001. Correlation modelling on the sphere using a generalized diffusion equation. Quarterly Journal of the Royal Meteorological Society 127, 1815–1846.

Yu, P., Kurapov, A.L., Egbert, G.D., Allen, J.S., Kosro, P.M., 2012. Variational assimilation of hf radar surface currents in a coastal ocean model off oregon. Ocean Modelling 49, 86–104.

Zavala-Garay, J., Wilkin, J., Arango, H., 2012. Predictability of mesoscale variability in the east australian current given strong-constraint data assimilation. Journal of physical oceanography 42, 1402–1420.

Zhang, W.G., Wilkin, J.L., Arango, H.G., 2010. Towards an integrated observation and modeling system in the new york bight using variational methods. part i: 4dvar data assimilation. Ocean Modelling 35, 119–133.