A NEW LAGRANGIAN BASED SHORT TERM PREDICTION METHODOLOGY FOR HF RADAR CURRENTS

Lohitzune Solabarrieta[1,2], Ismael Hernandez-Carrasco[3], Anna Rubio[2], Michael Campbell[1], Ganix Esnaola[4,5], Julien Mader[2], Burton H. Jones[1], Alejandro Orfila[3]

(1) KAUST, Red Sea Research Center, Integrated Ocean Processes, Saudi Arabia.
(2) AZTI Marine Research, Basque Research and Technology Alliance (BRTA), Pasaia, Spain
(3) Instituto Mediterráneo de Estudios Avanzados. IMEDEA (CSIC-UIB), 07190 Esporles, Spain.
(4) Nuclear Engineering and Fluid Mechanics Dept., UPV ,20018-Donostia, Spain.
(5) Joint Research Unit BEGIK, (IEO)- (UPV/EHU), 48620-Plentzia, Spain.

Corresponding author's email: lsolabarrieta@azti.es

ABSTRACT

The use of High Frequency Radar (HFR) data is increasing worldwide for different applications in the field of operational oceanography and data assimilation, as it provides real-time coastal surface currents at high temporal and spatial resolution. In this work, a Lagrangian based empirical real-time, Short-Term Prediction (L-STP) system is presented in order to provide short term forecasts of up to 48 hours of ocean currents. The method is based on finding historical analogues of Lagrangian trajectories obtained from HFR surface currents. Then, assuming that the present state will follow the same temporal evolution as the historical analogue did, we perform the forecast. The method is applied to two HFR systems covering two areas with different dynamical characteristics: the southeast Bay of Biscay and the central Red Sea. A comparison of the L-STP methodology with predictions based on persistence and reference fields are performed in order to quantify the error introduced by this approach. Furthermore, a sensitivity analysis has been addressed to determine the limit of applicability of the methodology regarding the temporal horizon of Lagrangian prediction. A real-time skill-score has been developed using the results of this analysis, which allows to identify periods when the short-term prediction performance is more likely to be low, and persistence can be used as a better predictor for the future currents.

1. INTRODUCTION

The coastal zone is under increasing human pressure. During recent decades coastal seas have been experiencing intensified activity for recreation, transport, fisheries and marine-related energy production, which, in many cases, results in serious damage to coastal marine ecosystems. A better understanding of the dynamical processes responsible for the surface oceanic transport is a prerequisite for the efficient management of the coastal ocean. Coastal processes are responsible for the transport and fate of multi-source pollutants like plastics, nutrients, jellyfish, harmful algal blooms, etc. Thus, improving the capacity of monitoring and forecasting the coastal area is key for the integrated assessment of the marine ecosystem. This requirement is driving the set-up of a growing number of multi-platform operational observatories designed for continuous monitoring of the coastal ocean from international or national (e.g., US IOOS, EU EOOS, Australian IMOS, etc.) to local scales. Moreover, due to the need of forecasting applications for response to emergency situations such as oil spills, or search and rescue operations, many of the existing operational observatories are linked with operational ocean forecasting models with or without data assimilation (e.g. MARACOOS, NOAA Global Real-Time Ocean Forecast System, COPERNICUS Marine Environment Monitoring System).

With the need of providing a long-term framework for the development and improvement of the European Marine coastal observations, the JERICO Research Infrastructure (JERICO-RI) has been developing methods and tools (through JERICO, JERICO-NEXT and JERICO-S3 projects) for the production of high-quality marine data, and sharing expertise and infrastructures between the existing observatories in Europe. Typically constituted with different in-situ point-wise observational platforms (such as moored buoys, tidal gauges, drifting buoys, etc.) a significant number of these observatories now employ land-based High Frequency Radars (HFR) that provide real-time coastal currents with unprecedented coverage and resolution (e.g. Paduan and Rosenfeld, 1996; Kohut and Glenn, 2003; Abascal et al., 2009; Solabarrieta et al., 2014, Rubio et al. 2017; Paduan and Washburn, 2013). Each HFR coastal site measures radial surface currents moving away or approaching the antenna, based on the shift of the first peak (Bragg peak) of the Doppler spectra (Crombie 1955, Barrick et al 1977). Combining the overlapping radial vectors from at least 2 antennas provides surface true vector currents (Barrick et al., 1977, Barrick and Lipa, 1979). Several studies have compared *in-situ* current measurements with HFR observations (e.g., Schott et al. 1985; Hammond et al. 1987; Paduan and Rosenfeld 1996, Emery et al. 2004; Paduan et al., 2006; Ohlmann et al. 2007; Liu et al., 2014; Solabarrieta et al, 2014, Bellomo et al., 2015; Lana et al., 2016; Hernandez-Carrasco et al., 2018b) and have repeatedly demonstrated the potential of this technology. Presently, more than 250 HFR antennas are installed and active worldwide (Roarty et al., 2019; http://global-hfradar.org/).

Due to their high spatio-temporal resolution, HFR data are commonly used in real time for search and rescue (Ullman et al., 2006) or oil spill prediction/mitigation emergency response (Abascal et al., 2017). In addition, there have been several efforts dedicated to the development of assimilation strategies that incorporate the HFR measured surface currents into ocean coastal models (Breivik and Saetra, 2001, Oke et al 2002, Paduan and Shulman 2004, Stanev et al., 2011, Barth et al., 2011), some of which have been tested for short periods of time (Chao et al., 2009). However, assimilation of HFR data into models is still a computationally expensive and complex issue, not to mention operational capabilities of such a procedure. Because of these constraints, the availability of real-time high-resolution HFR current fields has led to alternative solutions in order to obtain short term prediction (STP) of surface coastal currents, through the direct use of HFR historical and nowcast observations using different approaches (e.g. Zelenke 2005, Frolov et al. 2012, Barrick et al., 2012, Orfila et al. 2015, Solabarrieta et al. 2016, Vilibić et al, 2016, Ren et al., 2019, see Table 1).

The above-mentioned studies develop and implement different STP approaches (harmonic analysis of the last hours, genetic algorithms, numerical models, …) which often require either additional data, or long training periods of data without gaps. Hardware failures due to power issues, communications or environmental conditions often result in spatio-temporal gaps within HFR datasets. Spatial gaps can be filled on a real-time basis but filling long temporal gaps is not straightforward. Several gap-filling methodologies have been developed for HFR data sets: Open Modal Analysis, (OMA) (Kaplan and Lekien, 2007), Data Interpolating EOFs (DINEOF) (Hernandez-Carrasco et al., 2018), and Self-Organizing Maps (SOM) (Hernandez-Carrasco et al., 2018).

Given the motivation described above, and developed partially within the framework of JERICO-NEXT project, we present a Lagrangian-based Short-Term Prediction (L-STP from now on) methodology using existing HFR datasets, to be applied to surface current real-time observations. The proposed L-STP methodology aims to be capable to use the previously developed gap filling OMA method and generate forecasts in near-real time with low computational costs, compared to the previously presented forecast methods, but with the same level of assessment. The uniqueness of this approach is two-fold: first, the historical

Eulerian velocity fields are used to construct a catalogue of Lagrangian trajectories and second, using the trajectories obtained from present observations, analogues in the past dataset are searched in order to obtain the best predictive match. The method is based on Lagrangian computations which have proven to be robust against errors in velocity field data and against the dynamics of unresolved scales, since the averaging effect produced by integrating over trajectories which extend in time and space, tends to cancel random-like errors (Hernandez-Carrasco et al., 2011, Sayol et al., 2014). Consequently, they are reliable for the assessment of the dynamical flow structures.

Analogues is a widely used method in time series prediction, especially in early weather forecasting and statistical downscaling. It is based on the assumption that if the behavior of a dynamical system at a given time is similar or close enough to some other situation in the historical record, then the evolution in the future of the state of the system will be similar to the evolution observed in the same historical record. Simply stated, two analogue fields are two distinct fields that are close enough considering a given metric, to be considered as equivalent. Finding of the best (nearest) analogue of a specific time does not require a historically continuous dataset, as long as the dataset contains subsets of observations that extend longer than the testing period and are representative of the range of potential states that the system can have. These statistically analogue events occur naturally in the environment and this methodology has been applied and tested in atmospheric forecasts (Lorenz, 1969, Jianping et al,1993, Prince and Goswami 2007, Shao and Li 2013).

It must be stressed that this is the first time that the analogues technique has been applied to the HFR-derived ocean surface currents to obtain short-term forecast, to the knowledge of the authors. The L-STP is intended to be implemented operationally with low computational cost (seconds to few minutes for each forecast, depending on the size of the historical dataset) and is easily implemented using existing HFR data processing tools.

2. DATA AND METHODS

2.1 Data

HFR data from two distinct oceanographic regions have been used for the evaluation, validation, and testing of the developed methodology (Figure 1): The Bay of Biscay (hereinafter BoB HFR) and the central Red Sea region (hereinafter Red Sea HFR). The range and the spatial resolution of the HFR current systems depend on their working frequency and the conductivity of the water over which the system is measuring. Ranges vary from 15 to 220 km range and spatial resolution from 250 m to 12 km. Typically, a 12 MHz radar has a range ~70 km with a spatial resolution of 2-5 km. HFR systems usually average current measurements for one hour, although some average currents for shorter periods, such as 30 minutes. HFR data from these two regions are used to evaluate the skill of the method under different dynamical conditions, and with a sufficient set of observations to provide a database suited to the efficient research of appropriate analogues. The BoB HFR system, located in the southeastern corner of the Bay of Biscay, in the Basque Country, is composed of two CODAR Seasonde sites, working since 2009 at 4.5 MHz frequency, covering up to 200 km range and providing hourly surface velocity field at 5 km of spatial resolution. The dataset used in this study spans the period from January 2012 to December 2015. The Red Sea HFR system is located on the central western coast of Saudi Arabia and is also composed of two CODAR Seasonde sites. The Red Sea sites are operational since June 2017, transmit at 16.12MHz frequency, covering up to 120 km range and providing the hourly surface velocity field at 3 km spatial resolution. The dataset used in this study spans the period from June 2017 to October 2018.

The BoB HFR has been chosen as the pilot system for testing the developed methodology, since it has the longest data series and because several papers have already provided an extensive description of the local circulation and dynamical processes (Rubio et al., 2013a, 2013b, 2018, 2019, 2020; Solabarrieta et al 2014, Solabarrieta et al., 2015, Hernandez-Carrasco et al. 2018, Manso-Narvarte et al., 2018; Declerk et al., 2019). The resulting methodology is then applied to the operational Red Sea HFR dataset, as a study case. Coastal dynamics in the BoB show a clear seasonality where cyclonic and anticyclonic eddies dominate in winter and summer, respectively in responding to local winds and the mean coastal current (Iberian Poleward Current) (Esnaola et al., 2013, Solabarrieta et al., 2014). The circulation in the central Red Sea also demonstrates a clear seasonality (Sofianos and Johns, 2003; Yao et al., 2014a, 2014b; Zarokanellos et al., 2017a, 2017b) linked to the seasonal winds of the area (Abualnaja et al., 2015; Langodan et al., 2017b). The region is dominated by eddy activity, with both cyclonic and anticyclonic eddies occurring in the region (Zhan et al., 2014; Zarokanellos et al. 2017a). Due to the only recently available dataset (since mid-June 2017 to present) the detailed small-scale surface circulation processes of this area is under characterization at the moment.

The primary difference between the two HFR systems is the operating frequency resulting in a larger spatial coverage for the BoB HFR than for the Red Sea HFR and a higher spatial resolution for the latter (5km and 3 km, respectively). This difference in the spatial resolution should result in better capturing the small-scale dynamical features in the Red Sea, which could influence the selection of an analogue.

The data from both systems have been processed similarly. The spectra of the received backscattered signal are converted into radial velocities using the MUltiple SIgnal Classification (MUSIC) algorithm (Schmidt 1986). HFR Progs MATLAB package (https://github.com/rowg/hfrprogs) is then used to combine radial currents and generate gap-filled total 2D currents, by means of the Open Modal Analysis (OMA) methodology of Kaplan and Lekien (2007).

2.2 Lagrangian analogues

The proposed prediction system, based on the analogue identification method, has been developed with the objective of providing HFR velocity fields forecast (up to 48 hours). As an innovative element, we use a Lagrangian approach in searching for analogues through an historical library composed of particle trajectories, instead of the commonly used Eulerian velocity fields. In our methodology we find the best analogue by comparing maps of trajectories obtained from the last available 48 hours (target field) with the historical catalogue of maps of Lagrangian trajectories (hereinafter Lagrangian catalogue). Then the catalogue map with the trajectory pattern closest to the target field map is selected. Relying on the similar evolution of the current situation and the past analogue, the next 48-hour time velocity fields of the selected analogue provides the target period forecast. In other words, if we find a state in the historical database that is "close enough" to the target field, we assume that the forecast for the current observations will evolve in the same way as did for the chosen analogue. A detailed description of the short-term prediction system is provided in the following algorithm:

1. *Lagrangian catalogue configuration*. First, to build the Lagrangian catalogue, a set of synthetic trajectories was computed by advecting N particles uniformly initialized on a regular grid (Figure 2) in the OMA HFR velocity fields. The N Lagrangian particles are released every hour over the whole available velocity data and are advected during 48 hours. The maps of trajectories of the catalogue are referred as to $\mathbf{X}_C$.
2. *Target map*. A map of trajectories corresponding to the most recent HF currents observations, and referred as to $\mathbf{X}_T$, is computed using the same procedure than for the Lagrangian catalogue but now advecting the N particles in the available last 48 hours (t_f–48h) of HFR velocity fields, where t_f corresponds to the current time.

3. *Searching for the analogue.* A searching algorithm for the best (closest to the target map) analogue among all the trajectory maps is implemente next. To increase the efficiency of this process, the search was done in two steps.
   i. Optimization of the catalogue. First, selecting only "potential" analogues with a similar main drift reduces the Lagrangian catalogue. The trajectories centroid for each map of the catalogue is computed and compared to that of the target field, and finally discarding the analogues whose centroid was at a distance greater than δ_{cg}. The value of the δ_{cg} is selected to be small enough to minimize the computational time but sufficiently large to do not lose sampling variability in the potential analogues. We explored different values of this threshold distance to find that $\delta_{cg} = 2\xi = 10$ km (where ξ is the spatial resolution) makes a good compromise between computational cost and number of potential analogues in both study areas.
   ii. In a second step, we computed the Lagrangian errors (ε) between the trajectories of the target field and the potential analogues, defined as:

$$\varepsilon_{ANL} = \sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(\delta_{ANL}(t_i)\right)^2}, \quad t_i = \{6,12,24,36,48 \text{ hours}\}, \qquad (1)$$

where T = 5 is the number of elements of the set of times t_i, and $\delta_{ANL}(t_i)$ is the mean separation distance at time t_i between the trajectories belonging to the target field $\mathbf{X}_T$ and each of the potential analogues $\mathbf{X}_c$, given by:

$$\delta_{ANL}(t_i) = \frac{1}{N}\sum_{j=1}^{N} \left| \left(X_T^j(t_i) - X_c^j(t_i)\right) \right| , \qquad (2)$$

being N the total number of trajectories j.

4. *Best analogue.* The selection of the best analogue is performed by the Equation (2), which is a simple measure of similarity between two datasets. The best analogue is selected as the element of the catalogue with the lowest ε_{ANL}. Figure 3 shows an example of the time series of ε_{ANL} values, through the catalogue of potential analogues for a specific case. Then we locate the time t_{ANL} corresponding to best analogue: $t_{ANL} \rightarrow \min(\varepsilon_{ANL}) = \varepsilon_{ANL}(t_{ANL}) : X_c(t_{ANL})$.
5. *Currents Prediction.* Once we have identified t_{ANL}, the short term forecast of the HFR velocity fields is given by the hourly velocity fields corresponding to the next 48 hours since t_{ANL} (hereinafter "L-STP fields"):
$\mathbf{X}_{STP}(t_c+1:t_c+48h)=\mathbf{X}_c(t_{ANL}+1:t_{ANL}+48h)\rightarrow\mathbf{V}_{STP}(t_f+1:t_f+48h)=\mathbf{V}_c(t_{ANL}+1:t_{ANL}+48h)$,

where $\mathbf{V}_C$ (t_{ANL}) is the velocity field corresponding to the best analogue and $\mathbf{V}_{STP}$ are the forecast currents.

Figure 2 provides an example of the selected analogue (Figure 2b) and corresponding L-STP fields (Figure 2d) for a given target field (Figure 2a) and the 'truth' trajectories for the following 48 hours from the date of the target field (Figure 2c). The associated temporal series of errors for the target field and the potential analogues are shown in Figure 3, where the value of ε_{ANL} is marked using a red dot (corresponding to the error between the trajectories of the L-STP field in Figure 2d and the truth trajectories for the forecast period in –Figure 2c).

To assess the performance of the methodology, we computed forecasted trajectories based on persistence of currents (hereinafter 'persistence fields' $\mathbf{X}_{PRS}$). To obtain simulated trajectories using persistence currents, the particles are advected during 48 hours using a constant (frozen) velocity field (given by the current velocity field, or target field, $\mathbf{V}(t_f)$) during the 48 hours of simulation: $\mathbf{V}(x,y,t_f+T)= \mathbf{V}(x,y,t_f)$, where t_f = current time and T={1 : 48h}.

The mean drift of the truth forecasted trajectories, $\mathbf{X}_{TRU}$, is also computed for each simulation period (the mean drift is computed averaging over all the particle trajectory length during 48 hours).

The Lagrangian errors between the truth trajectories $\mathbf{X}_{TRU}$ and the L-STP trajectories $\mathbf{X}_{STP}$ were also computed as:

$$\varepsilon_{STP} = \sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(\delta_{STP}(t_i)\right)^2} \; = \sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\left(X_{TRU}^{j}(t_i) - X_{STP}^{j}(t_i)\right)\right)^2}\;, \qquad (3)$$

where δ_{STP} is the mean separation distance between truth and the L-STP trajectories for t= t : t+48 (following 48 hours from the study time). To compare with persistence, we also compute the Lagrangian error between the truth trajectories $\mathbf{X}_{TRU}$ and the trajectories derived from the persistence field X_{PRS},

$$\varepsilon_{PRS} = \sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(\delta_{PRS}(t_i)\right)^2} \; = \sqrt{\frac{1}{T}\sum_{j=1}^{T}\left(\frac{1}{N}\sum_{j=1}^{N}\left(X_{TRU}^{j}(t_i) - X_{PRS}^{j}(t_i)\right)\right)^2}\;,, \qquad (4)$$

where δ_{PRS} is the mean separation distance between truth maps of trajectories, X_{TRU}, and maps of trajectories from persistent velocity fields, $\mathbf{X}_{PRS}$, for t= t:t+48 (following 48 hours from the study time)

All the process for the selection and validation of the analogue with the different variables has been summarized in Figure 4. The time series and spatial distribution of the ε_{STP} and ε_{PRS} errors have been analyzed for both study areas. Finally, ε_{STP} and ε_{PRS} time series have also been calculated and compared to the time series of the ε_{ANL}, in order to evaluate if the ε_{ANL} can be used as an indicator of the expected skill of the L-STP with respect to the persistence.

Some parameters in the algorithm have to be tuned in order to optimize the results and the computational cost. For instance, we found that the optimal number of particle trajectories, N is equal to 25. All the trajectories have been computed considering infinitesimal and passive particles without adding a diffusion term. To this end we used the Lagrangian module included in the HFR_Progs MATLAB package.

The ability of this method relies on the precision in finding two matching HFR currents states over the entire region, which is dependant on the historical record of observations used to build the catalogue and the dynamical representativity of the catalogue. In this study we use four-year dataset (2012-2015) of trajectory maps computed for the SE BoB, where the trajectory maps from the three first years (2012-2014) were used as Lagrangian catalogue, and the remaining year (2015) was used as a test period. The historical Lagrangian catalogue for this HFR system is, thus, composed of 26304 maps of N=25 trajectories of 48-hours. Then the method was applied to the Red Sea dataset, for the period of July 2017-October 2018. As the dataset temporal extension was short (1 year and 4 months), we have used the whole period to build the Lagrangian catalogue and act as a test period at the same time. In this case, for the analogues search the 5-days period around the date of the target field was removed from the catalogue at each iteration, to avoid temporal overlapping with the target field.

3. RESULTS

Figure 2 shows an example of the developed methodology applied to the BoB HFR system on April 15, 2015. It is a visual representation of the (a) target trajectories, (b) the selected analogue, (c) truth trajectories during the next 48 hours from the target period, and (d) the L-STP trajectories provided by the method (48 hours from the analogue).

The performance assessment results are described in section 3.1 and the temporal and spatial forecast for both study areas are shown in section 3.2.

3.1 Assessment of the L-STP skills

Figure 5 shows the ε_{ANL} through 2015 for the BOB study area, together with the ε_{STP} and ε_{PRS}. The analysis of this plot aims to check the relation between $\varepsilon_{ANL,}$ ε_{STP} and ε_{PRS}. Black dots over the timeline in Figure 5 show the times when ε_{STP} is higher than the $\varepsilon_{PRS,}$ which occurs 12% of the time. The mean value of the ε_{PRS} is 73% higher than the ε_{STP}. The correlation between ε_{ANL} and ε_{STP} is 0.46 while correlation between ε_{ANL} and ε_{PRS} is 0.05, for the whole test year (2015). Focusing on the times when the ε_{PRS} is lower than the ε_{STP}, it can be seen that they mostly occur during winter months. Previous works in this area have shown that there are high persistent eastward currents that can last for several weeks during winter months (Solabarrieta et al., 2014), which can explain the better performance of the persistence fields in this period.

The hourly values of ε_{STP} and ε_{PRS} have been plotted against their corresponding hourly ε_{ANL} values for the test year, ordered from minimum to maximum along the x-axis in Figure 6. We observe that, when ε_{ANL} is low (less than 13.06 km for this data set), ε_{STP} is smaller than $\varepsilon_{PRS.}$ However, as ε_{ANL} increases, ε_{STP} and ε_{PRS} converge until an inflection point beyond which ε_{STP} is slightly greater than $\varepsilon_{PRS.}$ For the SE BoB experiment, the inflection point occurs at ε_{ANL} =13.06 km and 88% of cumulative ε_{ANL}. Results from the Red Sea HFR system indicates a similar pattern (not shown), when the inflection point occurs at ε_{ANL} = 12.81 km and at 86.4% of cumulative ε_{ANL}.

Further analysis to elucidate the mean separation distances (δ_{STP} and δ_{STP}) related to ε_{ANL} after 6, 12, 24, 36 and 48 hours are presented hereinafter. ε_{ANL} has been plotted together with the mean separation distances of the trajectories (δ_{STP} and δ_{PRS}), after 6, 12, 24, 36 and 48 hours for each target field (Figure 7). δ_{STP} is always higher than the δ_{PRS} for the 6 hours' simulation. But the values of δ_{STP} show lower values than δ_{PRS} for the lowest ε_{ANL} for the simulations at 12, 24, 36 and 48 hours.

The values of the correlation coefficient (R^2) between the ε_{ANL} and δ_{STP} and between ε_{ANL} and δ_{PRS} after 6, 12, 24, 36 and 48 hours are summarized in Table 2, in order to analyze the relations between the Analogue, the L-STP and the persistence. Values of R^2 for ε_{ANL} and δ_{PRS} are small (almost no correlation), varying between 0.01 and 0.11, while correlations between ε_{ANL} and δ_{STP} are higher, varying between 0.19 and 0.56, and showing higher correlation (>than 0.37) after 12 hours of simulations. The behavior of the Red Sea HFR system figures (not shown) is similar to the BoB HFR system.

Figures 6 and 7 (and the same ones for the Red Sea system, not shown) show that while ε_{ANL} increases, ε_{STP} and δ_{STP} increase, but ε_{PRS} and δ_{PRS} decrease, showing an inflexion point (hereinafter $\varepsilon_{ANL(*)}$). The $\varepsilon_{ANL(*)}$ can be calculated just for the historical dataset but ε_{ANL} can also be calculated in real time and compared with $\varepsilon_{ANL(*)}$. It gives a reference value for the forecast skills:

$\varepsilon_{ANL} < \varepsilon_{ANL(*)} \rightarrow \quad \delta_{\_STP} < \delta_{\_PRS} \quad \rightarrow$ Use L-STP

$\varepsilon_{ANL} > \varepsilon_{ANL(*)} \rightarrow \quad \delta_{\_STP} > \delta_{\_PRS} \quad \rightarrow$ Use Persistence

To assess the capabilities of the L-STP methodology, times when $\varepsilon_{ANL} < \varepsilon_{ANL(*)}$ have been just analyzed from now on, as when $\varepsilon_{ANL} > \varepsilon_{ANL(*)}$ we recommend to use persistent currents as a short term forecast.

3.2 Spatio-temporal performances of the L-STP methodology

Mean separation distances between truth and forecasted trajectories after different periods of integration times have been computed for both systems just for $\varepsilon_{ANL} < \varepsilon_{ANL(*)}$ times (Figure 6), in order to evaluate the temporal forecast capabilities of the methodology. Separation distances computed for the whole test year 2015, are shown in Figure 8, for the BoB HFR observations.

The separation distances between the measured trajectories and predicted persistent and STP trajectories, have similar values during the first 6 hours (4km) of the forecast period, with slightly better results for persistent trajectories. But after 6 hours, the separation distance for the forecast based on persistent currents increases faster than using L-STP. At 24 hours, the separation distance is 11 km for persistence forecasts and 8km for L-STP forecasts. The values are 12 and 18km, respectively, after 48 hours of simulation. The mean drift values of the truth trajectories show that the mean drift is similar to the L-STP separation distances, during the 48 hours.

Temporal mean separation distances between truth and forecasted trajectories for the Central Red Sea HFR System, computed for $\varepsilon_{ANL} < \varepsilon_{ANL(*)}$ are shown in Figure 9. The separation distances for the STP forecasts are higher than those forecasts with persistent currents during the first 15 hours. After 15 hours, quality of forecasts reversed where STP produced better results than persistence.

Spatial distribution of the difference between δ_{PRS} and δ_{STP} at 6, 12, 24 and 48 hours, for the BoB and the Red Sea study areas, are shown in Figure 10 and Figure 11.

For the BoB HFR system, the differences are not appreciated during the first 6 hours. However, after 12 hours of simulation, the advantage of the L-STP is clear in most of the study area, especially outside the continental shelf slope where persistent currents dominate the circulation. The separation values between δ_{PRS} and δ_{STP} increase up to 10km after 48hours of simulation.

For the Red Sea, the significant differences between STP and Persistence start after 24 hours of simulation, and continue until 48 hours.

4. DISCUSSION

In this work, a new methodology to forecast ocean surface currents based on HFR observations has been described. The approach is based on the search of analogues in a trajectory (Lagrangian) space using a previously generated trajectory field catalogue. The temporal and spatial skills of the proposed L-STP methodology have been analyzed in the previous section.

The target Lagrangian trajectory maps have been compared with the previously generated trajectory catalogue to obtain ε_{ANL}, ε_{STP}, ε_{PRS}, δ_{STP} and δ_{PRS} for each analyzed time. For the BoB system (2015 period), the correlation between ε_{ANL} and ε_{PRS} is 0.05, showing no relation between them and similar values are obtained for ε_{ANL} and δ_{PRS} (0.01-0.11- from table 2). The correlation between ε_{ANL} and ε_{STP} is 0.46 and it varies from 0.19 to 0.56 between ε_{ANL} and $\delta_{\_STP}$. Although the correlation between ε_{ANL} (past) and δ_{STP} or ε_{STP} (future) are low, they suggest that there is a relation between the errors of the analogues and the errors of the L-STP. δ_{STP} is always higher than the δ_{PRS} for the 6 hours' simulation. Which means that for the first hour, it is better to use persistence.

The $\varepsilon_{ANL(*)}$ can just be calculated for the historical dataset but ε_{ANL} can also be calculated and compared to the previously selected $\varepsilon_{ANL(*)}$, in real time. It gives a reference value for the forecast skills and we suggest that ε_{ANL} can be considered as a real-time skill-score metric for the L-STP :

$$\varepsilon_{ANL} < \varepsilon_{ANL(*)} \rightarrow \quad \delta_{\_STP} < \delta_{\_PRS} \quad \rightarrow \quad \text{Use L-STP}$$

$$\varepsilon_{ANL} > \varepsilon_{ANL(*)} \rightarrow \quad \delta_{\_STP} > \delta_{\_PRS} \quad \rightarrow \quad \text{Use Persistence}$$

The election of the best value for $\varepsilon_{ANL(*)}$ is the main sensitive step of the proposed methodology: the values of ε_{ANL} are different for each study area and no fixed value can be given. Due to this, an exhaustive analysis of $\varepsilon_{ANL,}$ δ_{STP}, δ_{PRS} of the historical dataset is required to find the correct inflexion point and select a correct $\varepsilon_{ANL(*)}$, before the method can be applied to a new study area.

Once fixed $\varepsilon_{ANL(*)}$, the skills of the proposed L-STP methodology have been tested in figures 8 to 11. The values of the δ_{STP}, compared to previous works in the BoB area showed that the L-STP produces accurate predictions, which demonstrates the ability of the Lagrangian approach to capture key dynamical features needed to accurately predict the proper dynamical conditions.

For the BoB HFR System, temporal δ_{STP} shows values of 3.5km, 5.5km and 8km, after 6, 12, and 24 hours respectively. The δ_{STP} values are similar to the δ_{PRS} values during the first 6 hours of simulation but δ_{STP} are lower after that, with 3km and 5.5km of difference between them, after 24 and 48 hours of simulation, respectively (*Figure* 8). As stated in previous work, that the circulation over the BoB area is dominated by a stable, persistent current field during winter (Solabarrieta et al., 2014) which is reflected by these results where persistence has good or even slightly better forecasting skill during the first 6 forecast hours than the proposed methodology.

δ_{STP} values for the BoB HFR system are similar to the ones obtained by Solabarrieta *et al.*, 2016, for the whole year but δ_{STP} are better for summer months, for the same study area. They used the linear autoregressive model, described in Frolov et al., 2012, to forecast HFR current fields and the errors using that approach were 2.9 and 7.9 km after 6 and 24 hours. Although the results obtained in this work improve only during certain periods the forecast presented in Solabarrieta et al., 2016, the presented methodology has three advantages over the previous method: it is easy to run in real time; it does not require a continuous training period; and it is able to discriminate the times when the usage of the persistence is applicable. On the negative side, it requires the generation of a catalogue of past trajectories as the search space for analogues, but once it is ready, it is easily increasable in real time, without extra pre-analysis; just adding new trajectory fields to the previous catalogue.

The values of the δ_{STP} for the Red Sea HFR system follow a similar pattern to the BoB results, with higher separation distances. This may be related to the limited time span of the available dataset, as a better closest analogue may be found in a longer dataset.

The spatial comparison of the δ_{STP} and δ_{PRS} for the BoB HFR system (Figure 10), shows that the L-STP has better skills for the entire study area after 12 hours of simulations. The skills of the L-STP with respect to the persistence increases with time, showing up to 10km of improvement relative to persistence at 48 hours in some parts of the study area. For the spatial distribution, after 12 hours, the smallest differences between δ_{STP} and δ_{PRS} occurred over the slope. This is explained by the existence of persistent seasonal Iberian Poleward Current that flows along the continental slope toward the east along the Spanish coast and northward along the French coast (Solabarrieta *et al.,* 2014). In other words: although the L-STP can be performant in periods of persistent currents, the persistence field can show a better forecast for a short temporal scale (48h). L-STP will improve those forecasts, as soon as spatio-temporal variability increases.

The results for the Red Sea HFR system are similar but the benefit of the L-STP methodology appears only after 12 hours of simulation. Spatially, the improvement is again lower where persistent currents occur, as it is the case of the Eastern Boundary Current that flows northward following the eastern Red Sea Coastline in the study area (Bower and Farrah, 2015; Sofianos and Johns, 2003; Zarokanellos *et al.,* 2017b). The dominance of the persistent currents is evident in the lower values of the difference between the STP forecasts and the Persistence forecasts as shown in Figure 11 and in comparison, with Figure 10.

We have compared the capabilities of the L-STP forecast against the forecast based on the persistency of currents. The L-STP method requires long training periods but performs better during non-persistent periods. Previous efforts to forecast surface currents from HFR data have shown similar results compared with the methodology presented in this paper. However, the advantage of the L-STP method is that it can be used in near real time, with short and non-continuous datasets of around 2-3 years, provided that a Lagrangian catalog representative for the study area can be built.

5. CONCLUSION

A methodology forecast surface currents with analogues of Lagrangian dynamics in real-time has been proposed. This methodology provides accurate forecast of sea surface currents up to 48 hours and its capability has been tested in terms of spatial and temporal distributions. The methodology has been successfully applied to two distinct coastal regions to evaluate its capabilities in different hydrodynamic regimes, although further analysis using data from more areas is required to generalize the methodology.

Relationships between ε_{ANL} and ε_{STP}/ ε_{PRS} suggest that the ε_{ANL} can be considered as a reliable indicator of the method's performance. Taking in consideration all the analyses done in this work, we propose to use STP currents for trajectory or velocity field predictions from 12 hours forward, if the ε_{ANL} value is lower than $\varepsilon_{ANL(*)}$. If ε_{ANL} is higher than $\varepsilon_{ANL(*)}$, or the forecast is just for the next 6 hours, the use of the persistence field is suggested. We also suggest that the $\varepsilon_{ANL(*)}$ value and forecast transition time need to be carefully evaluated for each study region. This, of course, infers that a minimum data set is required before the L-STP method can be applied.

Further analysis of analogue finding approaches is required to improve the observed results, especially during periods when currents are persistent. The use of longer dataset as a training period may improve this aspect. Then, the next step would be to test the methodology for additional periods and other regions, to analyze the possibility to find analogues for different sub-regions and to evaluate its functionality in an operational mode.

The methods to find the minimum training period for each system should be analyzed deeper in future works. The minimum training period will be directly related to the variability of the local dynamics and those should be considered during the analysis.

The HFR Progs MATLAB package (https://cencalarchive.org/~cocmpmb/COCMPwiki) has been used to generate total currents from radial files and to fill the spatial gaps of the surface current field using the OMA method, and to generate Lagrangian trajectories. The presented forecasting method can be therefore easily implemented as an additional tool to provide short term forecast at the same time that they generate total currents.

DATA AVAILABILITY

The Red Sea HF Radar data can be requested through:

- https://lthdatalib.kaust.edu.sa

Historical and NRT Bay of Biscay HF Radar data can be requested through:

- Euskoos portal: https://www.euskoos.eus/en/data/basque-ocean-meteorological-network/high-frequency-coastal-radars/
- Emodnet Physics - http://www.emodnetphysics.eu/Map/platinfo/piradar.aspx?platformid=10273
- CMEMS Instac - http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=INSITU_GLO_UV_NRT_OBSERVATIONS_013_048

AUTHOR CONTRIBUTION

- **Lohitzune Solabarrieta**: She has worked on the set up of the methodology, data analysis, manuscript writing and final submission.
- **Ismael Hernandez-Carrasco**: He has worked on the set up of the methodology and the manuscript writing.
- **Anna Rubio**: She has worked on the set up of the methodology, data analysis, and manuscript writing.
- **Michael Campbell**: He has worked on the configuration of the methodology. He has also contributed on the manuscript writing.
- **Ganix Esnaola**: He has worked on the configuration of the methodology. He has also contributed on the manuscript writing.
- **Julien Mader**: He has contributed on the writing of the manuscript.
- **Burton H. Jones**: He has contributed on the writing of the manuscript.
- **Alejandro Orfila**: He has worked on the configuration of the methodology, data analysis and the manuscript writing.

COMPETING INTERESTS

The authors declare that we have no conflict of interest

ACKNOWLEDGEMENTS

This work was funded by a Saudi Aramco-KAUST Center for Marine Environmental Observation (SAKMEO) Postdoc fellowship to Lohitzune Solabarrieta, and from the Integrated Ocean Processes (IOP) Group in KAUST. We acknowledge the support of the LIFE-LEMA project (LIFE15 ENV/ES/000252), the European Union's Horizon 2020 research and innovation program under grant agreement No. 654410 & 871153 (JERICO-NEXT and JERICO-S3 Projects), the Directorate of Emergency Attention and Meteorology of the Basque Government, the MINECO/FEDER Project MOCCA (256RTI2018-093941-B-C31). and the Department of Environment, Regional Planning, Agriculture and Fisheries of the Basque Government (Marco Program). This work was partially performed while A. Orfila was a visiting scientist at the Earth, Environmental and Planetary Sciences Department at Brown University through a Ministerio de Ciencia, Innovación y Universidades fellowship (PRX18/00218). Ismael Hernandez-Carrasco acknowledges the Vicenç Mut contract funded by the Balearic Island Govern and the European Social Fund (ESF) Operational Programme. The HF radar-processing toolbox HFR_Progs use to produce OMA was provided by D. Kaplan and M. Cook, Naval Postgraduate School, Monterey, CA, USA.

REFERENCES

Abascal, A. J., Castanedo, S., Medina, R., Losada, I. J., Álvarez-Fanjul, E.: Application of HF radar currents to oil spill modelling. Mar. Pollut. Bull. 58 (2), 238–248, 2009

Abascal A. J., Sanchez, J., Chiri, H., Ferrer, M. I., Cárdenas, M., Gallego, A., Castanedo, S., Medina, R., Alonso-Martirena, A., Berx, B., Turrell, W. R., Hughes, S. L.: Operational oil spill trajectory modelling using HF radar currents: A northwest European continental shelf case study. Marine Pollution Bulletin, Volume 119, Issue 1, Pages 336-350, ISSN 0025-326X, https://doi.org/10.1016/j.marpolbul.2017.04.010, 2017.

Abualnaja, Y., Papadopoulos, V. P., Josey, S. A., Hoteit, I., Kontoyiannis, H., and Raitsos, D. E.: Impacts of climate modes on air–sea heat exchange in the Red Sea, J. Clim., 28, 2665–2681, doi:10.1175/JCLI-D-14-00379.1, 2015.

Barrick, D. E.: Extraction of wave parameters from measured HF radar sea-echo Doppler spectra. Radio Sci., 12, 415–424, doi:10.1029/RS012i003p00415, 1977.

Barrick D.E., Fernandez, V., Ferrer, M.I., Whelan, C., and Breivik, Ø.: "A short-term predictive system for surface currents from a rapidly deployed coastal HF-Radar network," Ocean Dyn., vol. 62, no. 5, pp. 725–740, 2012.

Barth, A., Alvera-Azcárate, A., Beckers, JM., Staneva J., Stanev E.V., and Schulz-Stellenfleth J.: Correcting surface winds by assimilating high-frequency radar surface currents in the German Bight. Ocean Dynamics, 2011, vol 61: 599. https://doi.org/10.1007/s10236-010-0369-0, 2011.

Bellomo, L., Griffa, A., Cosoli, S., Falco, P., Gerin, R., Iermano, I., Kalampokis, A., Kokkini, Z., Lana, A., Magaldi, M.G., Mamoutos, I., Mantovani, C., Marmain, J., Potiris, E., Sayol, J.M., Barbin, Y., Berta, M., Borghini, M., Bussani, A., Corgnati, L., Dagneaux, Q., Gaggelli, J., Guterman, P., Mallarino, D., Mazzoldi, A., Molcard, A., Orfila, A., Poulain, P. M., Quentin, C., Tintoré, J., Uttieri, M., Vetrano, A, Zambianchi, E. and Zervakis, V.: Toward an integrated HF radar network in the Mediterranean Sea to improve search and rescue and oil spill response: the TOSCA project experience. 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, 8:2, 95-107, DOI: 10.1080/1755876X.2015.1087184, 2015.

Bower, A. S., and Farrar, J. T.: Air–sea interaction and horizontal circulation in the Red Sea. In N. M. A. Rasul & I. C. F. Stewart, (Eds.), The Red Sea, Springer Earth System Sciences (pp. 329–342). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-662-45201-1_19, 2015.

Breivik, Ø, and Saetra, Ø.: Real time assimilation of HF radar currents into a coastal ocean model. Journal of Marine Systems, Volume 28, Issues 3–4, April 2001, Pages 161-182. https://doi.org/10.1016/S0924-7963(01)00002-1, 2001.

Chao, Y., Li Z., Farrara, K., McWilliams, J.C., Bellingham, J., Capet, X., Chavez, F., Choi, J., Davis, R., Doyle, J., Fratantoni, D. M., Li P., Marchesiello, P., Moline, M.A., Paduan, J., Ramp, S.: Development, implementation and evaluation of a data-assimilative ocean forecasting system off the central California coast. Deep Sea Research, Vol. 56, Issues 3-5, pp 100-126. https://doi.org/10.1016/j.dsr2.2008.08.011, 2009.

Crombie, D. D.: Dopler Spectrum of Sea Echo at 13.56-Mc/s’, Nature 175, 681-682, 1955.

Declerck A., Delpey M., Rubio A., Ferrer L., Basurko O. C., Mader J. and Louzao M. Transport of floating marine litter in the coastal area of the south-eastern Bay of Biscay: A Lagrangian approach using modelling and observations, Journal of Operational Oceanography, 12:sup2, S111-S125, DOI: 10.1080/1755876X.2019.1611708, 2019.

Emery, B. M., Washburn L., and Harlan, J. A.: Evaluating radial current measurements from CODAR high-frequency radars with moored current meters. J. Atmos. Oceanic Tech- nol., 21, 1259–1271, doi:10.1175/1520-0426(2004)021,1259: ERCMFC.2.0.CO;2, 2004.

Esnaola, G., Sáenz, J., Zorita, E., Fontán, A., Valencia, V., and Lazure, P.: Daily scale wintertime sea surface temperature and IPC-Navidad variability in the southern Bay of Biscay from 1981 to 2010, Ocean Sci., 9, 655–679, https://doi.org/10.5194/os-9-655-2013, 2013.

Frolov, S., Paduan J., Cook M., and Bellingham J.: Improved statistical prediction of surface currents based on historic HF- radar observations. Ocean Dyn., 62, 1111–1122, doi:10.1007/ s10236-012-0553-5, 2012.

Hammond, T.M., Pattiaratchi ,C.B., Osborne, M.J., Nash, L.A., Collins, M.B.: Ocean surface current radar (OSCR) vector measurements on the inner continental shelf. Continental Shelf Research. Volume 7, Issue 4, Pages 411-431. https://doi.org/10.1016/0278-4343(87)90108-7, 1987.

Hernández-Carrasco, I., López, C., Hernández-García, E. & Turiel, A. How reliable are finite-size Lyapunov exponents for the assessment of ocean dynamics? Ocean Modelling 36, 208–218, 2011.

Hernández-Carrasco, I., Solabarrieta, L., Rubio, A., Esnaola, G., Reyes, E., and Orfila, A.: Impact of HF radar current gap-filling methodologies on the Lagrangian assessment of coastal dynamics, Ocean Sci., 14, 827-847, https://doi.org/10.5194/os-14-827-2018, 2018.

Hernández-Carrasco, I., Orfila, A., Rossi, V., and Garçon, V.: Effect of small-scale transport processes on phytoplankton distribution in coastal seas, Scientific Reports, 8:8613, https://doi.org/10.1038/s41598-018-26857-9, 2018b.

Jianping H., Yuhong Y., Shaowy W and Jifen C.: An analogue-dynamical long-range numerical weather prediction system incorporating historical evolution. Q. J. R. Meteorol. Soc., 119, pp.547-565, 1993.

Kaplan, D. M. and Lekien, F.: Spatial interpolation and filtering of surface current data based on open-boundary modal analysis, Journal of Geophysical Research: Oceans, 112, https://doi.org/10.1029/2006JC003984, c12007, 2007.

Kohut, J.T., Glenn, S.M.: Improving HF radar surface current measurements with measured antenna beam patterns. J. Atmos. Oceanic Technol. 20 (9), 1303–1316, 2003.

Lana, A., Marmain, J., Fernández, V., Tintoré, J., Orfila A.: Wind influence on surface current variability in the Ibiza Channel from HF Radar. Ocean Dynamics, 66: 483. https://doi.org/10.1007/s10236-016-0929-z, 2016.

Langodan S., Cavaleri L., Vishwanadhapalli Y., Pomaro A., Bertotti L., Hoteit I.:. Climatology of the Red Sea - Part 1: the wind. Int. J. Climatol. 37: 4509–4517. DOI: 10.1002/joc.5103, 2017.

Liu Y., Weisberg, R.H., and Merz, C.R.: Assessment of CODAR SeaSonde and WERA HF Radars in Mapping Surface Currents on the West Florida Shelf. Journal of atmospheric and oceanic technology, Vol 31., pp 1363:1382, 2014.

Barrick D.E. and Lipa B.J. (1979) A Compact Transportable HF Radar System for Directional Coastal Wave Field Measurements. In: Earle M.D., Malahoff A. (eds) Ocean Wave Climate. Marine Science, vol 8. Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-3399-9_7, 1979.

Lorenz, E. N.: Atmospheric Predictability as Revealed by Naturally Ocurring Analogues. Journal of Atmospheric Sciences, Volume 29, pp 636-646, 1969.

Manso-Narvarte, I., Caballero, A., Rubio, A., Dufau, C., and Birol, F.: Joint analysis of coastal altimetry and high-frequency (HF) radar data: observability of seasonal and mesoscale ocean dynamics in the Bay of Biscay, Ocean Sci., 14, 1265–1281, https://doi.org/10.5194/os-14-1265-2018, 2018.

Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D., and Kosro P. M.: Assimilation of surface velocity data into a primitive equation coastal ocean model, J. Geophys. Res., 107, 3122, doi:10.1029/2000JC000511, 2002.

Orfila A., Molcard, A., Sayol, J.M., Marmain, J., Bellomo, L., Quentin, C., and Barbin, Y.: Empirical Forecasting of HF-Radar Velocity Using Genetic Algorithms IEEE Transactions on Geoscience and Remote Sensing, Vol. 53, No. 5, 2015.

Ohlmann, C., White, P., Washburn, L., Emery, B., Terrill, E., Otero, M.: Interpretation of coastal HF radar–derived surface currents with high-resolution drifter data. J. Atmos. Oceanic Technol. 24 (4), 666–680, 2007.

Paduan, J.D., and Rosenfeld, L.K.: Remotely sensed surface currents in Monterey Bay from shore-based HF radar (coastal ocean dynamics application radar. J. Geophys. Res. 101 (C9), 20669–20686, 1996.

Paduan, J.D., and Shulman, I.: HF radar data assimilation in the Monterey Bay area. J. Geophys Res. 109:C07S09, 2004.

Paduan, J.D., Kim, K.C., Cook, M. S., and Chavez, F.P.: Calibration and Validation of Direction-Finding High-Frequency Radar Ocean Surface Current Observations. IEEE Journal of oceanic engineering, Vol. 31, No. 4, 2006.

Paduan, J.D., and Washburn, L.: High-Frequency Radar Observations of Ocean Surface Currents. Annual Rev. Marine. Sci. 2013.5:115-136, 2013.

Prince, K., X. and Goswami, B., N.: An Analog Method for Real-Time Forecasting of Summer Monsoon Subseasonal Variability. Monthly weather review, Vol 135, pp: 4149-4160. https://doi.org/10.1175/2007MWR1854.1, 2007.

Ren L., Miaro, J., Li Y., Luo, X., Li J. and Hartnett, M.: Estimation of Coastal Currents Using a Soft Computing Method: A Case Study in Galway Bay, Ireland. Mar. Sci. Eng., 7(5), 157; https://doi.org/10.3390/jmse7050157, 2019.

Roarty, H., Cook, T., Hazard, L., George, D., Harlan, J., Cosoli, S., Wyatt, L., Alvarez Fanjul, E., Terrill, E., Otero, M., Largier, J., Glenn, S., Ebuchi, N., Whitehouse, B., Bartlett, K., Mader, J., Rubio, A., Corgnati, L., Mantovani, C., Griffa, A., Reyes, E., Lorente, P., Flores-Vidal, X., Saavedra-Matta, K.J., Rogowski, P., Prukpitikul, S., Lee, S.H., Lai, J.W., Guerin, C.A., Sanchez, J., Hansen, B. and Grilli, S.: The Global High Frequency Radar Network. Front. Mar. Sci. 6:164. doi: 10.3389/fmars.2019.00164, 2019.

Rubio, A., Fontán A., Lazure P., González M., Valencia V., Ferrer L., Mader J. and Hernández C., 2013a: Seasonal to tidal variability of currents and temperature in waters of the continental slope, southeastern Bay of Biscay. *J. Mar. Syst.*, **109–110** (Suppl.), S121–S133, doi:10.1016/j.jmarsys.2012.01.004.

Rubio, A., Solabarrieta L. , González M., Mader J., Castanedo S., Medina R., Charria G. and Aranda J., 2013b: Surface circulation and Lagrangian transport in the SE Bay of Biscay from HF radar data. *MTS/IEEE OCEANS—Bergen, 2013*, IEEE, 7 pp., doi:10.1109/OCEANS-Bergen.2013.6608039

Rubio, A., Mader, J., Corgnati, L., Mantovani, C., Griffa, A., Novellino, A., Quentin, C., Wyatt, L., Schulz-Stellenfleth, J., Horstmann, J., Lorente, P., Zambianchi, E., Hartnett, M., Fernandes, C., Zervakis, V., Gorringe, P., Melet, A., and Puillat, I.: HF radar activity in European coastal seas: next steps towards a pan-European HF radar network, Front. Mar. Sci., 4, 8, https://doi.org/10.3389/fmars.2017.00008, 2017.

Rubio, A., Caballero, A., Orfila, A., Hernández-Carrasco, I., Ferrer, L., González, M., Solabarrieta, L., and Mader, J.: Eddy-induced cross-shelf export of high Chl-a coastal waters in the SE Bay of Biscay, Remote Sens. Environ., 205, 290–304, 2018.

Rubio, A., Manso-Narvarte, I., Caballero, A., Corgnati, L., Mantovani, C., Reyes, E., et al. The seasonal intensification of the slope iberian poleward current. copernicus marine service ocean state report. J. Oper. Oceanogr. 12, 13–18. doi: 10.1080/1755876X.2019.1633075, 2019.

Rubio, A., Hernández-Carrasco, I., Orfila, A., González, M., Reyes, E., Corgnati, L., et al. (2020). A lagrangian approach to monitor local particle retention conditions in coastal areas. copernicus marine service ocean state report. J. Oper. Oceanogr. 13:1785097, 2020.

Sayol, J.M., Orfila, A., Simarro, G., Conti, G., Renault, L., Molcard, A. A Lagrangian model for tracking surface spills and SaR operations in the ocean, Env. Mod. & Software, (52), 74-82, 2014. doi:10.1016/j.envsoft.2013.10.013.

Schmidt, R.: Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag., 34, 276–280, doi:10.1109/TAP.1986.1143830, 1986.

Schott F., Frisch, A.S., Leaman, K., Samuels, G., Popa Fotino, I.: High-Frequency Doppler Radar Measurementsof the Florida Current in Summer 1983. J. Geo. Research, Vol 90, No C5, pp 9006:9016, 1985.

Shao, Q. and Li, M.: An improved statistical analogue downscaling procedure for seasonal precipitation forecast . Stoch Environ Res Risk Assess 27, pp.: 819-830. https://doi.org/10.1007/s00477-012-0610-0, 2013.

Sofianos, S. S., and Johns, W. E.: An oceanic general circulation model (OGCM) investigation of the Red Sea circulation: 2. Three- dimensional circulation in the Red Sea. Journal of Geophysical Research, 108(C3), 3066. https://doi.org/10.1029/2001jc001185, 2003.

Solabarrieta, L., Rubio, A., Castanedo, S., Medina, R., Charria, G., Hernández, C.: Surface water circulation patterns in the southeastern Bay of Biscay: new evidences from HF radar data. Cont Shelf Res 74:60–76 doi:10.1016/j.csr.2013.11.022, 2014.

Solabarrieta, L., Rubio, A., Cárdenas, M., Castanedo, S., Esnaola, G., Méndez, F.J., Medina, R., and Ferrer, L.: Probabilistic relationships be- tween wind and surface water circulation patterns in the SE Bay of Biscay. Ocean Dyn., 65, 1289–1303, doi:10.1007/s10236-015-0871-5, 2015.

Solabarrieta, L., Frolov, S., Cook, M., Paduan, J., Rubio, A., González, M., Mader, J., and Charria, G.: Skill Assessment of HF Radar–Derived Products for Lagrangian Simulations in the Bay of Biscay. J. Atmos. Oceanic Technol., 33, 2585–2597, doi: 10.1175/JTECH-D-16-0045.1, 2016.

Stanev, E.V., Schulz-Stellenfleth, J., Staneva, J., Grayek, S., Seemann, J. and Petersen, W.: Coastal observing and forecasting system for the German Bight – estimas of hydrophysical states. Ocean Sci., 7, 569–583, 2011 doi:10.5194/os-7-569-2011, 2011.

Ullman, D.S., O'Donnell, J., Kohut, J., Fake, T., Allen, A.: Trajectory prediction using HF radar surface currents: Monte Carlo simulations of prediction uncertainties. J. Geophys. Res. 111 (C12005), 1–14, 2006.

Vilibić, I., Šepić, J., Mihanović, H., Kalinić, H., Cosoli, S., Janeković, I., Žagar, N., Jesenko, B., Tudor, M., Dadić, V. dnd Ivanković, D.: Self-organizing maps-based ocean currents forecasting system. Scientific Reports 6, 22924, 2016.

Yao, F., Hoteit, I., Pratt, L. J., Bower, A. S., Zhai, P., Kohl, A., and Gopalakrishnan, G.: Seasonal overturning circulation in the Red Sea: 1. Model validation and summer circulation, J. Geophys. Res. Oceans, 119, doi:10.1002/ 2013JC009004, 2014a.

Yao, F., Hoteit, I., Pratt, L. J., Bower, A. S., Kohl, A., Gopalakrishnan, G., and Rivas, D.: Seasonal overturning circulation in the Red Sea: 2. Winter circulation, J. Geophys. Res. Oceans, 119, 2263–2289, doi:10.1002/ 2013JC009331, 2014b.

Zarokanellos, N. D., Kürten, B., Churchill, J. H., Roder, C., Voolstra, C. R., Abualnaja, Y., and Jones, B. H.: Physical mechanisms routing nutrients in the central Red Sea. Journal of Geophysical Research: Oceans, 122. https://doi.org/10.1002/2017JC013017, 2017a.

Zarokanellos, N. D., Papadopoulos, V. P., Sofianos, S. S., and Jones, B. H.: Physical and biological characteristics of the winter-summer transition in the Central Red Sea. Journal of Geophysical Research: Oceans, 122, 6355–6370. https://doi.org/10.1002/2017JC012882, 2017b.

Zhan, P., Subramanian, A. C., Yao, F., and Hoteit, I.: Eddies in the Red Sea: A statistical and dynamical study, J. Geophys. Res. Oceans, 119, 3909–3925, doi:10.1002/2013JC009563, 2014.

Zelenke B. C.: An Empirical Statistical Model Relating Winds and Ocean Surface Currents. Master of Science in Oceanography - Thesis, Oregon State University, 2005.

TABLES

Table 1*: Characteristics of the previously developed STP works based on HFR data.*

Authors	**Approach**	**Needs continuous training period**	**Comple-mentary data required?**	**Region of application**	**Reliable forecast period**
Zelenke 2005	EOF + bilinear regression model	Yes	Wind	Oregon coast	48 hours
Frolov et al. 2012	EOF + linear auto regression model	Yes	Wind and tides (optional)	Monterey Bay, California	48 hours
Barrick et al., 2012	Constant linear trend model applied to OMA modes	Yes	Wind	Finnmark, Norway	12 hours
Orfila et al. 2015	EOF+Genetic Algorithm	Yes	No	Toulon, France	48 hours
Solabarrieta et al. 2016	Frolov et al., 2012	Yes	No	Bay of Biscay	48 hours
Vilibić et al., 2016	SOM+neural network +winds	Yes	Wind	Northern Adriatic Sea	72 h
Ren et al., 2019	Random Forest (RF) classification algorithm	No	Tide and Wind	Galway Bay, Ireland	59 h
This paper: L-STP	Analogue finding	No	No	Bay of Biscay and the Central Red Sea	48 h

Table 2: Correlation coefficient values between best ε_{ANL} *and* δ_{STP} *and between* ε_{ANL} *and* δ_{PRS}*, after 6, 12, 24, 36 and 48 hours of simulation.*

	6 hours	**12 hours**	**24 hours**	**36 hours**	**48 hours**
$\mathbf{R^2}$ $\boldsymbol{\varepsilon_{ANL} - \delta\_{STP}}$	0.19	0.37	0.55	0.56	0.54
$\mathbf{R^2}$ $\boldsymbol{\varepsilon_{ANL} - \delta\_{PRS}}$	0.07	0.11	0.03	0.01	0.04
ε_{ANL} [km], for the inflection point between δ_{STP} and δ_{PRS}	-	11.94	12.44	13.09	14.33
% of ε_{ANL} (accumulative) for the previous line	-	81	84	87	95

FIGURES

Figure 1*: (Up) A globar view of both analyzed study areas. (Down-Left) HFR system of the BoB. (Down-Right) HFR system of the central Red Sea. Blue dots represent the data points and the black cross are the HFR antenna positions*

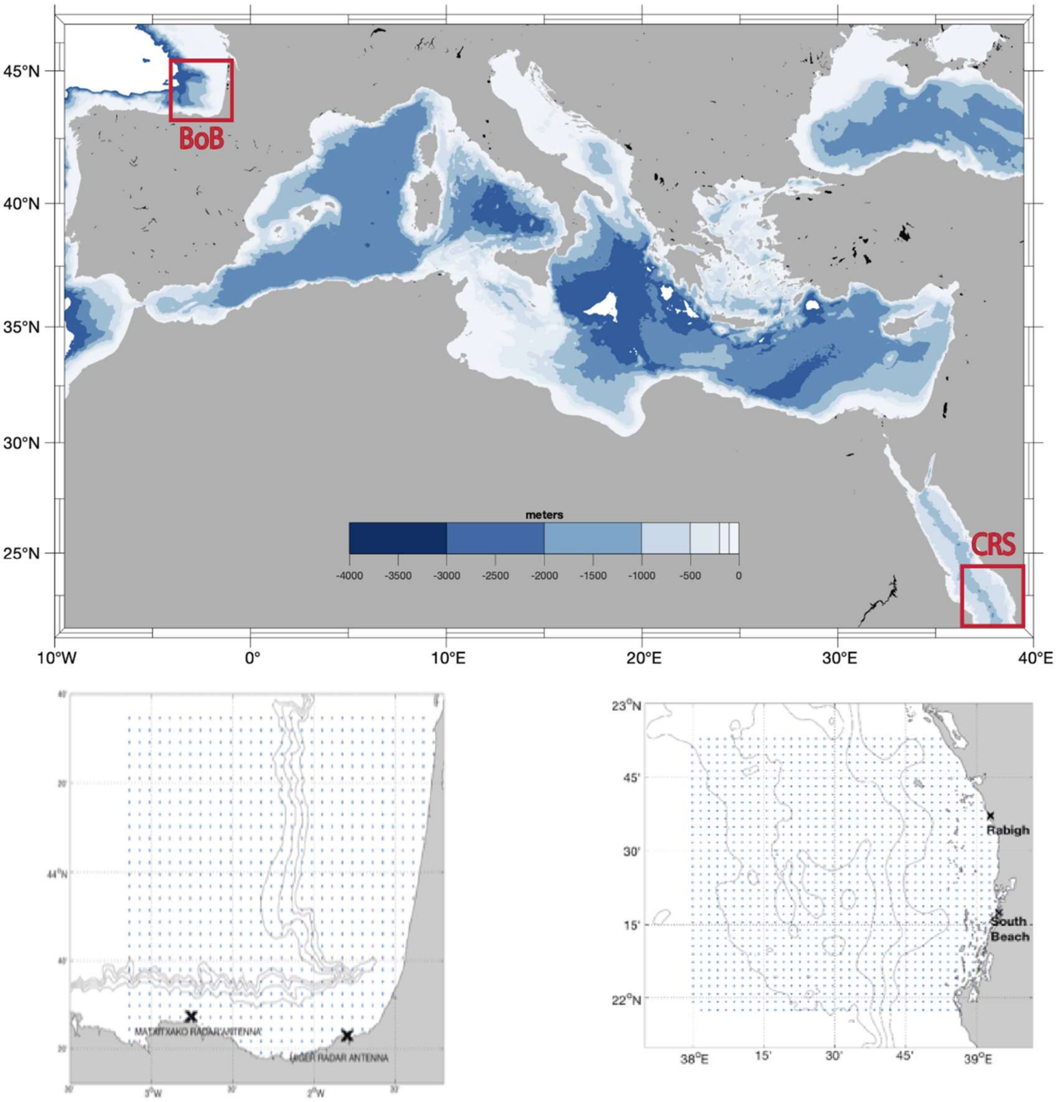


Figure 2: (1) 15-Apr-2015 00:00 example of the developed methodology applied to the BoB HFR system. (a) The past 48 hours of target field of test period (b) The analogue having the lowest error, (c) The truth trajectories for the forecast period (d) the STP trajectories. The initial positions of the particle trajectories are indicated by the blue dots, and the red dots indicate the position after 48 hours.

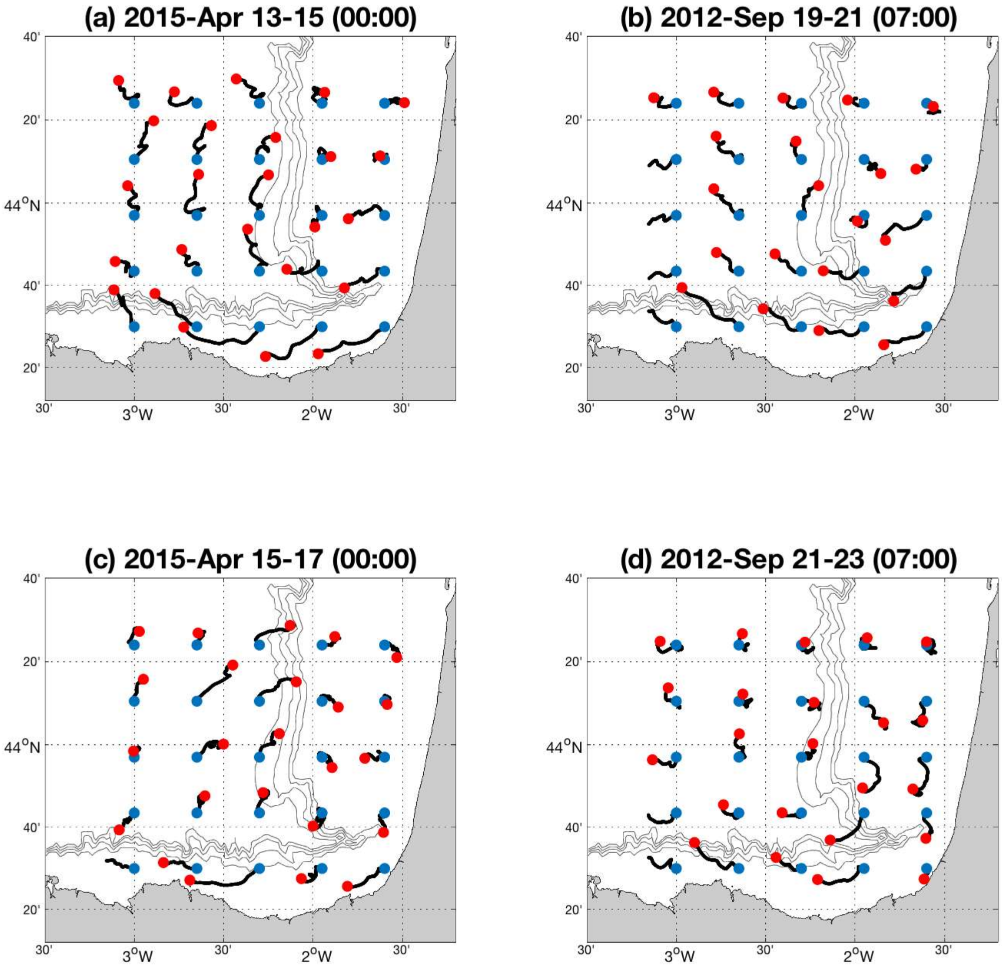


Figure 3***:*** *Example for the test period: 15-Apr-2015 00:00; errors for the whole Lagrangian catalogue fields of the BoB HFR System, restricted to the δ_cg = 10 km condition. The red dot indicates the occurrence date and the error of the best analogue (19-Sep-2012 07:00).*

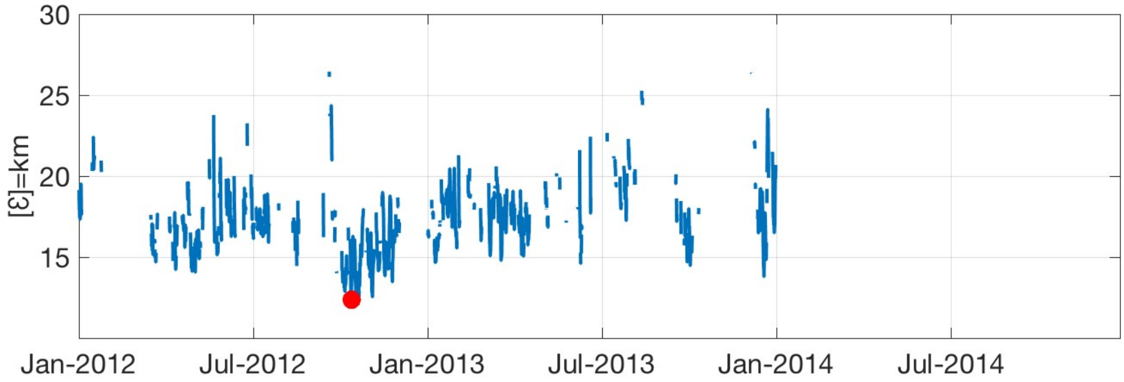


Figure 4: Scheme of the analogue selection and L-STP forecast assessment process.

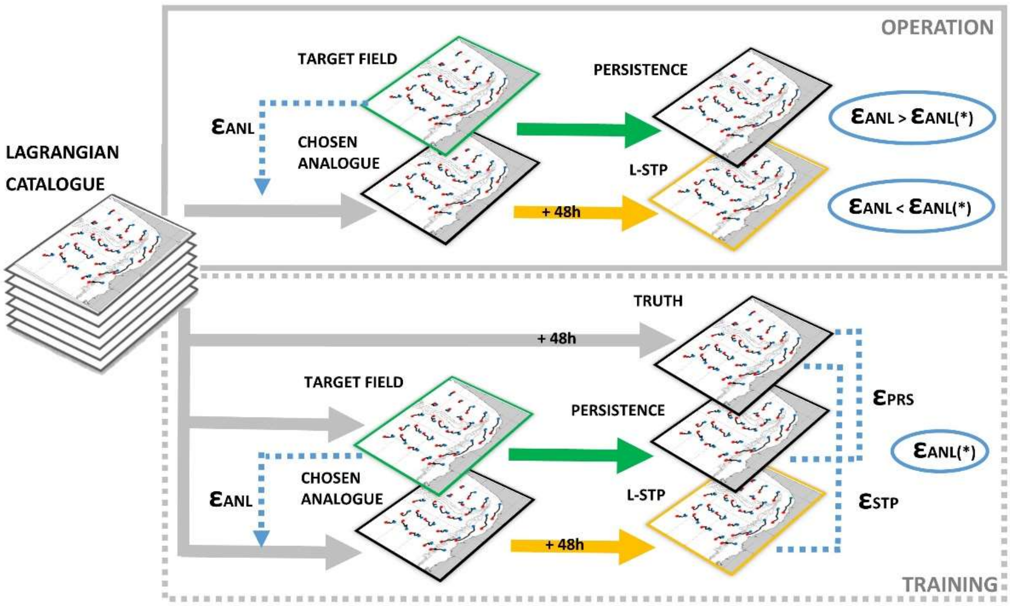


Figure 5: errors of the hourly best analogue for the BoB HFR, for 2015 (ε_{ANL}), together with the ε_{STP} and ε_{PRS}. The black dots over the timeline show the times when ε_{STP} is higher than ε_{PRS}

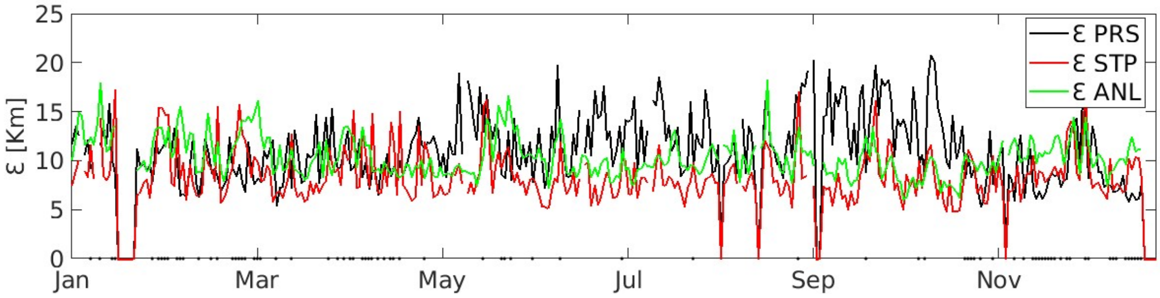
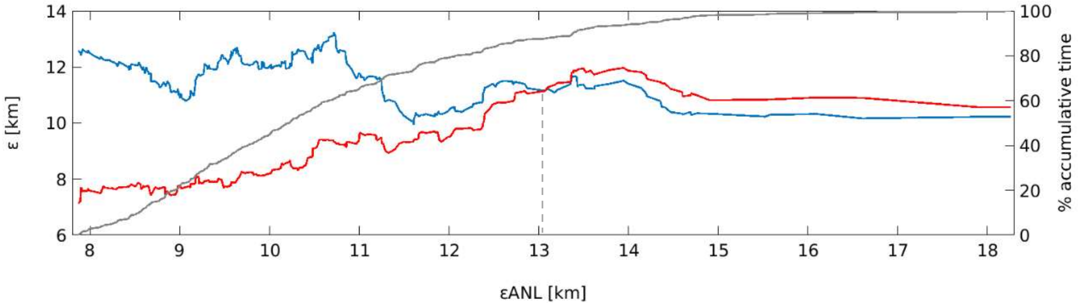


Figure 6: X axis shows the ε_{ANL}, ordered from minimum to maximum, for the best analogue for the test year 2015, for the BoB HFR. Left Y axis indicates ε_{STP} (red) and ε_{PRS} (blue) for the corresponding ε_{ANL}. Right Y axis indicates the % of the accumulative comparison times as shown by the gray solid line. Dashed vertical line indicates the crossing point between ε_{STP} and ε_{PRS} (ε_{ANL*}=13.06 km).



837 *Figure 7: Left Y axis indicates $\delta_{\_STP}$ (red) and $\delta_{\_PRS}$ (blue) for the corresponding ε_{ANL}, after 6, 12, 24, 36 and 48 hours. Right Y axis is the cumulative % of timesteps*
839 *in the computation of the mean errors, as indicated by the black line in the plots. X*
840 *axis is the ε_{ANL}, ordered from minimum to maximum, for the best analogue for the*
841 *test year 2015 (BoB HFR system)*

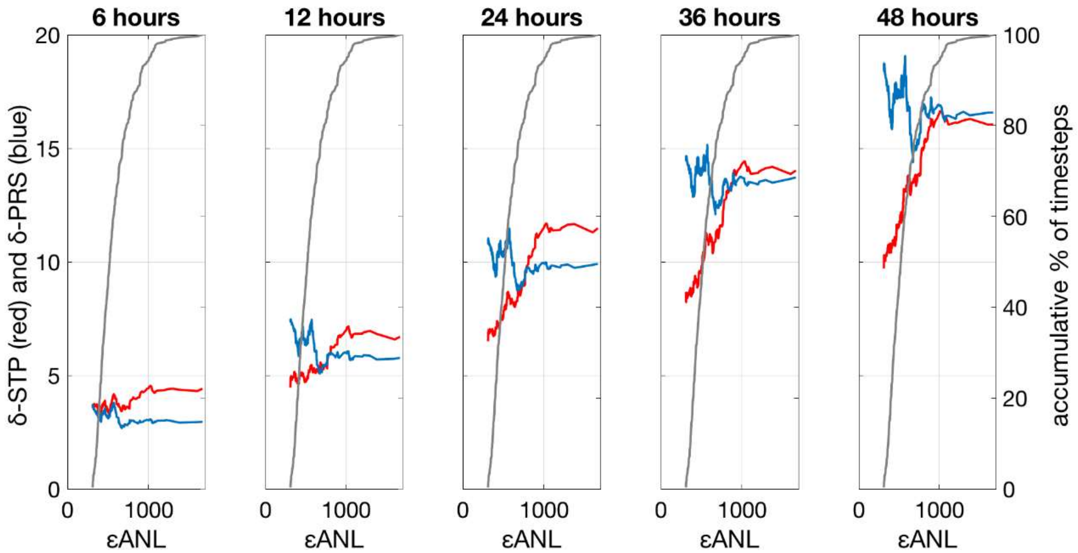
842 

Figure 8***:*** *Time evolution of the mean separation $\delta_{\_STP}$ and $\delta_{\_PRS}$ [km] between truth and forecast trajectories using truth and STP/PRS currents and the mean drift, with BoB system data, for 2015.* The mean drift of the truth forecasted trajectories is also computed for each simulation period (the means drift is considered as the average of the distances moved by each particle during 48 hours).

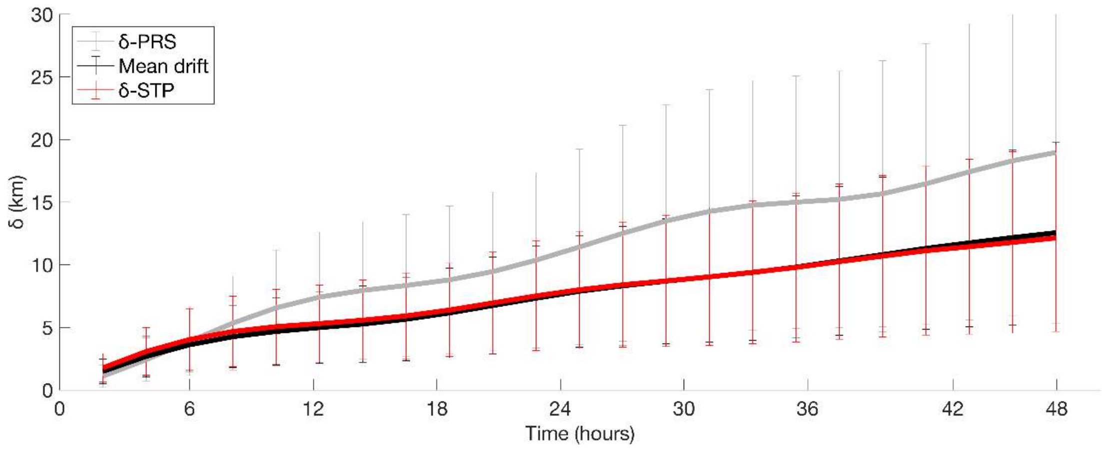


Figure 9***:*** *Time evolution of the mean separation distances $\delta_{\_STP}$ and $\delta_{\_PRS}$ [km] between real and forecast trajectories using truth and STP/PRS currents and the mean drift, with the Red Sea HFR system data, for July 2017 to October 2018.* The mean drift of the truth forecasted trajectories is also computed for each simulation period (the means drift is considered as the average of the distances moved by each particle during 48 hours).

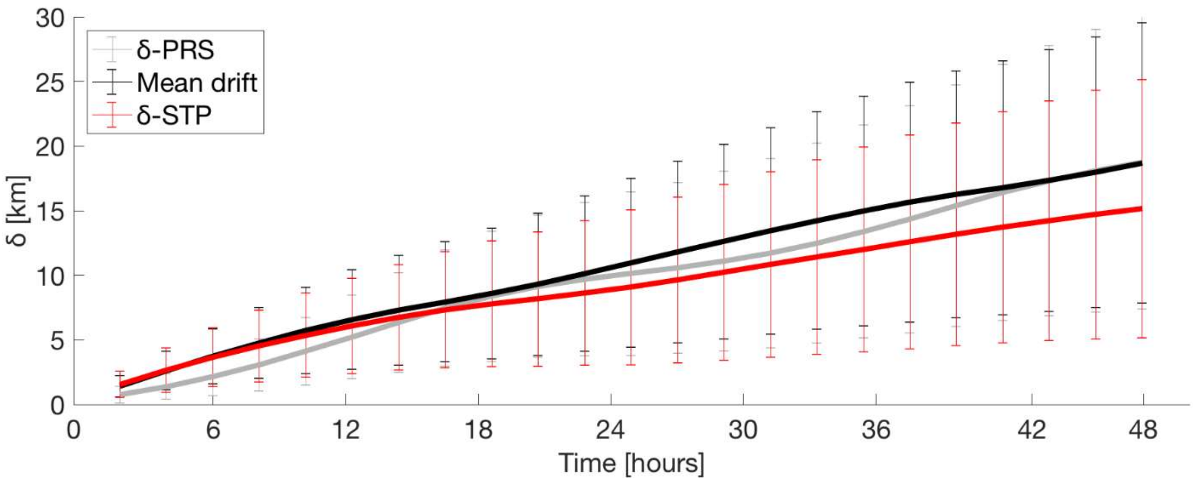


Figure 10***:*** *Spatial distribution of separation distances [km] between trajectories using L-STP and persistent currents at 6, 12, 24 and 48 hours, for the BoB HFR System.*

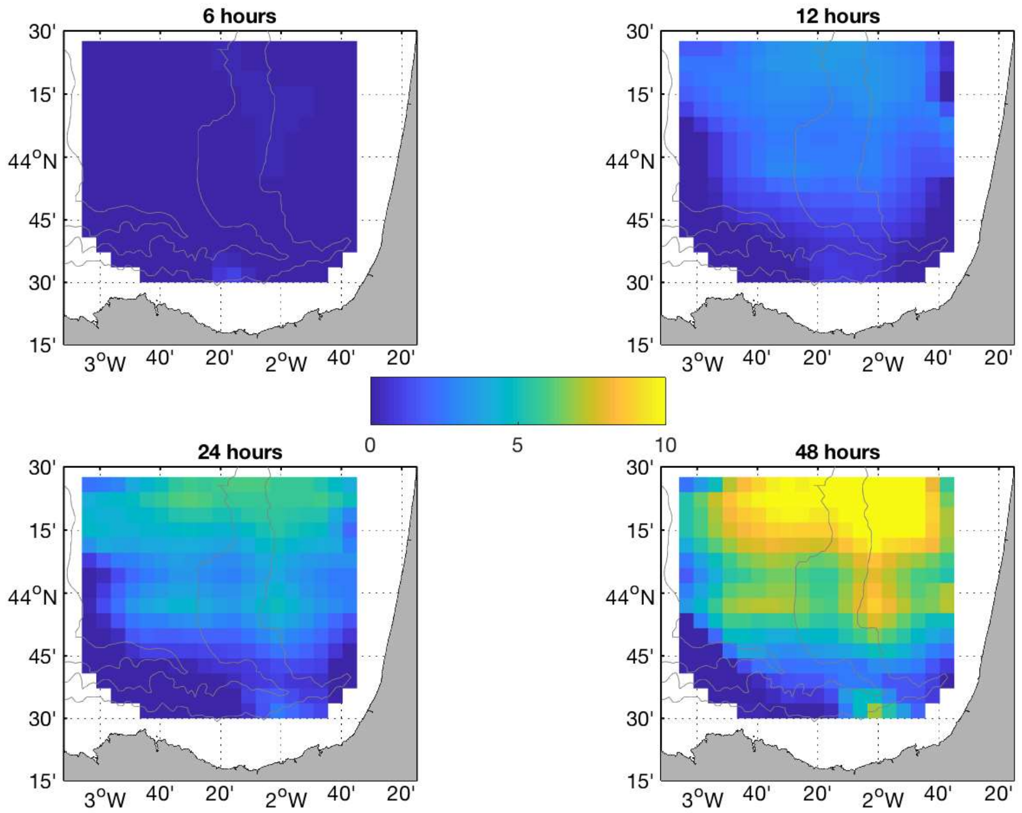


Figure 11: Spatial distribution of separation distances [km] between trajectories using L-STP and persistent currents at 6, 12, 24 and 48 hours, for the Red Sea HFR system.