Universitat de les Illes Balears

DOCTORAL THESIS
2022

IMPACTO DE LA ASIMILACIÓN DE OBSERVACIONES COSTERAS EN UN MODELO DE PREDICCIÓN OCEÁNICO DE ALTA RESOLUCIÓN.

IMPACT OF THE ASSIMILATION OF COASTAL OBSERVATIONS ON HIGH RESOLUTION OCEAN MODEL FORECASTS

Jaime Hernández Lasheras

Universitat de les Illes Balears

DOCTORAL THESIS
2022

Doctoral Programme in Physics

IMPACTO DE LA ASIMILACIÓN DE OBSERVACIONES COSTERAS EN UN MODELO DE PREDICCIÓN OCEÁNICO DE ALTA RESOLUCIÓN.

IMPACT OF THE ASSIMILATION OF COASTAL OBSERVATIONS ON HIGH RESOLUTION OCEAN MODEL FORECASTS

Jaime Hernández Lasheras

Thesis Supervisor: Baptiste Mourre
Thesis Supervisor: Alejandro Orfila Förster
Thesis Tutor: Damià Gomis Bosch

Doctor by the Universitat de les Illes Balears

Abstract

Operational Oceanography can be defined as the systematic routine collection of oceanic data and its interpretation for decision making. This research field has greatly evolved in the last decades, due to the exponential growth of the computational capabilities and the development of a global ocean observing system, which includes global remote sensed observations from satellites, multiple in-situ platforms, and high-resolution observations from newly developed coastal platforms. Within this context, the objective of this Thesis is to evaluate the impact of new coastal observing datasets in a high-resolution ocean circulation model, the WMOP (Western Mediterranean Operational System), using data assimilation (DA). DA is a powerful approach to merge observations and models in an optimal way to forecast ocean circulation as realistically as possible. A Local Multimodel Ensemble Optimal Interpolation data assimilation scheme is embedded in the WMOP system and configured to be able to ingest both large scale data all over the domain and high-resolution local observations. The study focuses on the Western Mediterranean Sea, an important biological and economical region exposed to strong anthropogenic pressure, where reliable forecasting models are crucial for the management of the coastal systems.

Three different experiments are developed, where the impact of glider fleet observations, dense CTD survey and high-frequency radars (HFR) measurements is evaluated.

First, we compare the performance of the DA system when using CTD versus a fleet of 8 gliders sampling a coastal area during the REP14 campaign, west of Sardinia. Several simulations assimilating CTDs or different number of gliders, in addition to generic observing sources (satellite sea level anomalies, surface temperature, and Argo profiles) are presented. The simulations were evaluated using independent data from CTD casts and a Scanfish gathered during the last part of the campaign. Results show that the assimilation of generic observing sources only, helps to represent the observed ocean

state better. Adding high-resolution local data from the campaign further increases the performance. The error between model and observations decreases as we increment the number of gliders sampling the area, with the best performance achieved when using eight-gliders, with a 40% error reduction. These results are similar to the ones obtained with the assimilation of 10 km spaced CTDs. Glider platforms, which can operate in all weather conditions, or in regions inaccessible to research vessels, and with a reduced cost, are demonstrated to be a very good alternative to the traditional ship-based campaigns.

Next, we evaluated the capacity of HFR observations to correct surface currents in the Ibiza Channel. Six different simulations were run, exploring different datasets and initialization methods after the analysis. In particular, the performance of whether using radial or total HFR observations together with generic observing sources is investigated. A set of 13 drifters deployed in the region are used for an independent validation, comparing the mean distance between the floats and the trajectory of virtual particles generated from the different simulations. The assimilation of satellite sea level anomalies, surface temperature, and Argo profiles helps to better represent the currents in the area. This performance is further improved when including HFR observations. The best results are obtained using reconstructed total current observations, reducing by 53% the average separation distance between drifters and virtual particles after the first 48 hours of simulation compared to the control run without assimilation.

Finally, an Observing System Simulation Experiment (OSSE) is performed to evaluate the potential impact of two future antennas in the Ibiza Channel HFR system. The two antennas would expand the actual coverage, providing surface currents observations in the whole channel. In the OSSE framework, a Nature Run (NR) simulation is used to represent the real ocean state, and pseudo-observations are extracted from it. To validate this framework, we first perform an OSSE using the same observation dataset as the real observation experiment from the previous Chapter. The pseudo-observations generated and the assessment of the impact on the simulations are consistent with the real observations experiment. The impact of the two new antennas is then evaluated in two one-month-long periods with different dynamical conditions. The impact is relatively small with the typical flow regime, due to the similarity of the NR and control run. In specific periods where the simulations present larger differences, expanding the

coverage results in a 19% error reduction with respect to the use of HFR observations from the actual system. Furthermore, we demonstrate that the assimilation of HFR observations helps to better recreate the Lagrangian Coherent Structures present in the NR, improving the representation of the ocean dynamics and the transport processes in the area.

Overall, in this Thesis we have demonstrated the importance of combining new high resolution coastal observing systems with traditional ones, to help constrain the circulation in a regional ocean model and enhance its forecast capabilities.

Resumen

La Oceanografía operacional se puede definir como la recolección continua y sistemática de observaciones en el océano para su tratamiento, análisis e interpretación para la toma de decisiones. Este campo de investigación ha evolucionado enormemente en las últimas décadas, gracias al crecimiento exponencial de las capacidades computacionales y al desarrollo del sistema global de observaciones, que incluye datos de teledetección provenientes de satélites, diferentes plataformas in-situ y nuevas fuentes de observación de alta resolución en zonas costeras. El objetivo principal de esta tesis es evaluar el impacto de los datos medidos por nuevos sistemas de observación costeros en un modelo regional de circulación oceánica mediante asimilación de datos. La asimilación de datos son un conjunto de técnicas estadísticas que pretenden emplear de forma óptima la información procedente de modelos y observaciones para mejorar las predicciones y hacerlas lo más realistas posibles. El modelo empleado es el WMOP (*Western Mediterranean Operational System*), el cual tiene incorporado un esquema secuencial de asimilación de datos (*Local Multimodel Ensemble Optimal Interpolation*) configurado para ser capaz de asimilar tanto observaciones de larga escala en todo el dominio, como datos de alta resolución en una única parte de este. El estudio está centrado en el Mediterráneo occidental, una región de un importante valor biológico y económico, expuesta a una gran presión antropogénica, lo que enfatiza la necesidad de desarrollar modelos de predicción operacionales fiables que ayuden en la gestión sostenible de las costas y mares.

A lo largo de esta tesis se presentan tres experimentos distintos en los que se evalúa el impacto de observaciones provenientes de planeadores submarinos (gliders), muestreos CTD y corrientes superficiales medidas con radares de alta frecuencia (HFR, por sus siglas en inglés).

Primero se compara el impacto en el modelo entre asimilar datos de CTD o de una

flota de *gliders*, aprovechando a campaña oceanográfica REP14, desarrollada en 2014 frente a la costa oeste de Cerdeña. Durante la campaña 8 *gliders* muestrearon la zona viajando en paralelo, a la vez que dos buques realizaban estaciones de medida de perfiles de CTD en el mismo área. Aquí se presentan los resultados de distintas simulaciones en las que se asimilan datos o bien de CTD o bien de distinto número de *gliders*, en ambos casos junto con datos genéricos (altimetría y de temperatura superficial provenientes de satélite y perfiles de boyas Argo). Las simulaciones han sido evaluadas con datos independientes de CTD y Scanfish, recogidos durante la última parte de la campaña. Los resultados muestran como, con la asimilación únicamente de los datos genéricos se consigue una mejoría en las predicciones y que esta mejoría es aún mayor cuando se emplean, además, datos de la campaña. Se observó que el error entre las observaciones y el modelo disminuye a medida que utilizamos más *gliders* para muestrear la zona, logrando hasta un 40% de reducción del error al emplear 8 *gliders*. Estos resultados son muy similares a los obtenidos al asimilar los perfiles CTD, equiespaciados 10 km. Se demuestra, por tanto, que los *gliders* son una gran alternativa a las tradicionales campañas en buques, al poder operar bajo cualquier estado de mar y en zonas inaccesibles para los barcos, reduciendo además los costes.

A continuación, se evalúa la capacidad las observaciones de HFR para corregir la predicción de las corrientes superficiales en el canal de Ibiza. Se presentan seis simulaciones distintas, en las que se emplean distintos tipos de observaciones y diferentes métodos de reinicialización del modelo después del análisis. En concreto, se evalúa la diferencia entre emplear observaciones radiales o totales (u-v) de HFR, ambas asimiladas junto con datos genéricos. Empleamos 13 boyas de deriva (*drifters*) lanzadas en la zona para validar el experimento, comparando la distancia promedio entre las boyas y partículas virtuales, que generamos para cada simulación. La representación de las corrientes superficiales en la zona mejora ya solamente con la utilización de observaciones genéricas y es aún mejor cuando incluimos, además, medidas de HFR. Los mejores resultados son obtenidos para la simulación que emplea observaciones totales, disminuyendo en un 53% la distancia promedio entre *drifters* y boyas virtuales transcurridas 48 horas.

Finalmente se evalúa el impacto que supondría la instalación de dos nuevas antenas en el canal de Ibiza. Las nuevas antenas en el lado occidental del canal ampliarían

la cobertura actual, proporcionando así medidas de corrientes superficiales en todo el canal. Ello se realiza mediante lo que se conoce como un experimento de simulación de sistema de observación (OSSE, en inglés). En el marco OSSE se emplea una simulación llamada *Nature Run* (NR), la cual es considerada como una representación realista del océano y que emplearemos para simular pseudo-observaciones. Para validar el marco OSSE primero realizamos un experimento idéntico al experimento real del capítulo anterior, pero usando observaciones virtuales. Tanto las pseudo-observaciones generadas, como el impacto de estas en el modelo, son consistentes con los resultados del experimento previo de referencia. El impacto de las nuevas antenas es evaluado en dos periodos de tiempo distintos, con diferentes condiciones dinámicas. Bajo el régimen típico de circulación, el impacto es relativamente bajo, debido a la semejanza que presenta el NR y la simulación de control. Sin embargo, cuando la diferencia entre estas simulaciones es más evidente, el empleo de observaciones del futuro sistema HFR puede llegar a disminuir el error de las predicciones en un 19%, comparado con el uso únicamente de datos de las antenas actuales. Además, en este capítulo demostramos como la asimilación de datos puede ayudar a recrear las estructuras coherentes Lagrangianas (LCS) presentes en el NR, mejorando así las dinámicas y los procesos de transporte en la zona.

En general, en esta tesis se demuestra la importancia de combinar sistemas de observación de datos costeros de alta resolución con fuentes de observación más tradicionales, ayudando a mejorar las predicciones de los modelos regionales.

Resum

L'Oceanografia operacional es pot definir com la recol·lecció contínua i sistemàtica d'observacions en l'oceà per al seu tractament, anàlisi i interpretació per a la presa de decisions. Aquest camp de recerca ha experimentant una gran evolució en les últimes dècades, gràcies al creixement exponencial de les capacitats computacionals i al desenvolupament del sistema global d'observacions, que inclou dades de teledetecció provinents de satèl·lits, diferents plataformes in-situ i noves fonts d'observació d'alta resolució en zones costaneres. L'objectiu principal d'aquesta tesi és avaluar l'impacte de les dades mesurades per nous sistemes d'observació costaners en un model regional de circulació oceànica mitjançant assimilació de dades, per a millorar les seves prediccions. L'assimilació de dades són és el un conjunt de tècniques estadístiques que pretenen emprar de manera òptima la informació procedent de models i observacions per tal dea millorar les prediccions i fer-les el més realistes possibles. El model emprat és el WMOP (*Western Mediterranean Operational System*), que incorpora un esquema seqüencial d'assimilació de dades (*Local Multimodel Ensemble Optimal Interpolation*) configurat per a ser capaç d'assimilar tant observacions de llarga escala en tot el domini, com a dades d'alta resolució en una única part d'aquest. El nostre estudi està centrat en el Mediterrani occidental, regió d'un important gran valor biològic i econòmic, sotmesa a una gran pressió antropogènica, la qual cosa emfatitza la necessitat de desenvolupar models de predicció operacionals fiables que ajudin en la gestió sostenible de les costes i mars.

Al llarg d'aquesta tesi es presenten tres experiments diferents en els quals s'avalua l'impacte d'observacions provinents de planadors submarins (*gliders*), mostrejos CTD i corrents superficials mesures amb radars d'alta freqüència (HFR, per les seves sigles en anglès).

Primer es compara l'impacte en el model entre assimilar dades de CTD o d'una

flota de *gliders*, aprofitant a campanya oceanogràfica REP14, desenvolupada en 2014 al litoral oest de Sardenya. Durant la campanya 8 *gliders* van mostrejar la zona viatjant en paral·lel, alhora que dos vaixells realitzaven estacions de mesura de perfils de CTD en la mateixa àrea. Aquí es presenten els resultats de diferents simulacions en les quals s'assimilen dades o bé de CTD o bé de diferent número de *gliders*, en tots dos casos juntament amb dades genèriques (altimetria i de temperatura superficial provinents de satèl·lit i perfils de boies Argo). Les simulacions han estat avaluades amb dades independents de CTD i Scanfish, recollits durant l'última part de la campanya. Els resultats mostren com, amb l'assimilació únicament de les dades genèriques s'aconsegueix una millora en les prediccions i que aquesta millora és encara major quan s'empren, a més, dades de la campanya. Es va observar que l'error entre les observacions i el model disminueix a mesura que utilitzem més *gliders* per a mostrejar la zona, aconseguint fins a un 40% de reducció de l'error en emprar 8 *gliders*. Aquests resultats són molt similars als obtinguts en assimilar els perfils CTD, espaiats cada 10 km. Es demostra, per tant, que els *gliders* són una gran alternativa a les tradicionals campanyes en vaixells, en poder operar sota qualsevol estat de mar i en zones inaccessibles per als vaixells, reduint a més els costos.

A continuació, s'avalua la capacitat les observacions de HFR per a corregir la predicció dels corrents superficials en el canal d'Eivissa. Es presenten sis simulacions diferents, en les quals s'empren diferents tipus d'observacions i diferents mètodes de reinicialització del model després de l'anàlisi. En concret, s'avalua la diferència entre emprar observacions radials o totals (u-v) de HFR, totes dues assimilades juntament amb dades genèriques. Emprem 13 boies de deriva (*drifters*) llançades en la zona per a validar l'experiment, comparant la distancia mitjana entre les boies i partícules virtuals, que generem per a cada simulació. La representació dels corrents superficials en la zona millora ja solament amb la utilització d'observacions genèriques i és encara millor quan incloem, a més, mesures de HFR. Els millors resultats són obtinguts per a la simulació que empra observacions totals, disminuint en un 53% la distancia mitjana entre *drifters* i boies virtuals transcorregudes 48 hores.

Finalment s'avalua l'impacte que suposaria la instal·lació de dues noves antenes en el canal d'Eivissa. Les noves antenes en el costat occidental del canal ampliarien la cobertura actual, proporcionant així mesurades de corrents superficials en tot el canal.

Això es realitza mitjançant el que es coneix com un experiment de simulació de sistema d'observació (OSSE, en anglès). En el marc OSSE s'empra una simulació anomenada *Nature Run* (NR), la qual és considerada com una representació realista de l'oceà i que emprarem per a simular pseudo-observacions. Per a validar el marc OSSE primer realitzem un experiment idèntic a l'experiment real del capítol anterior, però usant observacions virtuals. Tant les pseudo-observacions generades, com l'impacte d'aquestes en el model, són consistents amb els resultats de l'experiment previ de referència. L'impacte de les noves antenes és avaluat en dos períodes de temps diferents, amb diferents condicions dinàmiques. Sota el règim típic de circulació, l'impacte és relativament baix, a causa de la semblança que presenta el NR i la simulació de control. No obstant això, quan la diferència entre aquestes simulacions és més evident, l'ús d'observacions del futur sistema HFR pot arribar a disminuir l'error de les prediccions en un 19%, comparat amb l'ús únicament de dades de les antenes actuals. A més, en aquest capítol vam demostrar com l'assimilació de dades pot ajudar a recrear les estructures coherents Lagrangianes (LCS) presents en el NR, millorant així les dinàmiques i els processos de transport en la zona.

En general, en aquesta tesi es demostra la importància de combinar sistemes d'observació de dades costaneres d'alta resolució amb fonts d'observació més tradicionals, ajudant a millorar les prediccions dels models regionals.

Contents

Chapter 1

Introduction

1.1 Context

The ocean is a critical component of the Earth System, regulating weather, climate, nutrient distribution, the concentration of greenhouse gases, and redistributing energy across the globe. Research and general interest in ocean science have increased in the last years, and recently the United Nations has proclaimed a Decade of Ocean Science for Sustainable Development (2021–2030) to tackle the scientific challenges that are necessary for the sustainable use of natural resources (Pascual and Macías, 2021). Before that, oceans have attracted people's attention since centuries ago. Many ancient human civilizations arose near big rivers, which provided fresh water, food supplies, and fluvial transport means, but people have been colonizing coastal areas since very soon. Especially around the Mediterranean coasts, many civilizations established on the coasts and used the seas to feed themselves and trade with other places. Since then, the ocean has fascinated, scared and interested peoples living by their sides, as revealed by ancient texts and myths. Oceanography as a science did not arise until the end of the 19th century with the Challenger Expedition carried between 1872 and 1876. This circumnavigation is considered the first scientific expedition, with a proper dedicated oceanographic ship with its own equipment and laboratories, gathering a wide range of ocean characteristics including ocean temperature, seawater chemistry, currents, marine life and the seaflor geology.

However, it was not until the 20th century where technological advances permitted a fast growth in ocean research and in particular in operational oceanography which is the branch of the oceanography providing systematic and long-term routine measurements of the ocean and atmosphere and their rapid interpretation and dissemination (Chassignet, 2021). The evidence of the degradation of the oceans and coastal areas by pollution, unsustainable fishing practices, acidification, and temperature rise is nowadays overwhelming. These stressors represent a real threat resulting in an unprecedented decline in marine biodiversity, enhancing the importance of studying the ocean towards science-based management of our resources.

The dynamical processes in the ocean cover a wide range of scales. At the large scale mean currents and permanent structures with characteristic size of more than 500 km and periods between months and years play a central role in climate and climate variability by storing and transporting heat, fresh water and carbon around the globe. Below, the mesoscale (order of hundred of km) and submesoscale (order of km) dynamics dominate the main variability of the ocean (Chelton et al., 2007). These processes play a major role in redistributing properties and significantly impacting the ocean's

primary productivity. Assuming that the primary energy source for ocean currents occurs at the planetary scale, global energy flux equilibrium requires that this energy input is balanced by viscous dissipation at the smallest scales of motion (Pearson and Fox-Kemper, 2018). In the ocean, such scales are O(1 cm) or less. Gyre-scale circulations spawn energetic fields of mesoscale eddies (spatial scale 10's of kilometers, time scales of weeks). At these scales, motions are strongly constrained by rotation and stratification.

Satellite altimetry has revolutionized the understanding of ocean surface circulation during the last two decades (e.g., Le Traon (2013)), leading to major breakthroughs, as the quantification of Eddy Kinetic Energy (Escudier et al., 2016b), or Eddy identification and tracking (Chelton et al., 2011; Mason et al., 2014). However, altimetry still lacks enough resolution to cover scales shorter than 200 km, typical of ubiquitous mesoscale and submesoscale features (fronts, meanders, eddies, and filaments). Instead, it is necessary to rely on new advances and integrated approaches using high-resolution autonomous, in-situ and remotely sensed observations, modeling and fundamental processes studies (Hernández-Carrasco et al., 2018). All of these components are key elements of today's ocean forecasting systems (Schiller et al., 2018).

Marine observation systems, data management and dissemination have experienced a fast growth during the last decade (Sloyan et al., 2019; Moltmann et al., 2019; deYoung et al., 2019; Tintoré et al., 2019). Oceanographic information, combined with integrated predictive models, are increasingly needed to efficiently manage and effectively mitigate the eventual effects that global change are impinging on the seas; to portray the state of the ocean today, next week and for the next decade; to increase the efficiency of shipping; to mitigate storm damage and flooding of coastal areas; to sustain fisheries; to protect important ecosystems from degradation; to develop science-based sustainable management of marine and coastal areas and to improve climate forecasting. However, the ocean changes continuously, and therefore, the ocean must be monitored regularly to deliver accurate and reliable ocean services.

1.2 The Ocean Observing System. From Global to regional data

Sea Level Anomaly

Operational oceanography critically depends on near-real-time observations. The global Earth observing system has completely changed since the launch of satellites in the late 80s. Satellite altimetry is an essential observation required for global operational oceanography, and its implementation is understood as the first revolution in the observation of the ocean (Le Traon, 2013). A timeline of altimetry missions can be seen in Figure 1.1 below.

Radar altimeters measure sea surface height (SSH) relative to a given earth ellipsoid (Le Traon et al., 2019). A set of different satellite missions gather measurements every 7 km along repetitive tracks ranging between 10 days (TOPEX/Poseidon and Jason series) to 35 days (ERS, ENVISAT and SARAL/Alti-Ka) and a distance between tracks inversely proportional to the repetition period (between 315 and 90 km at the Equator). The altimeter constellation provides global, high spatial and temporal resolution measurements, not dependant on weather. At mid-latitudes, the sea level is, to a good approximation, an integral measure of the density variations of the upper ocean interior (Schiller et al., 2018). Geostrophic currents can be derived from SSH allowing

the identification of mesoscale structures and the main currents driving the ocean. Future SWOT mission, which is intended to be launched in 2022, will be based on new swath measurements, providing 2D SSH data on a 1-2 km grid (Morrow et al., 2018). It will suppose a revolution increasing the effective resolution and starting to capture even sub-mesoscale structures.

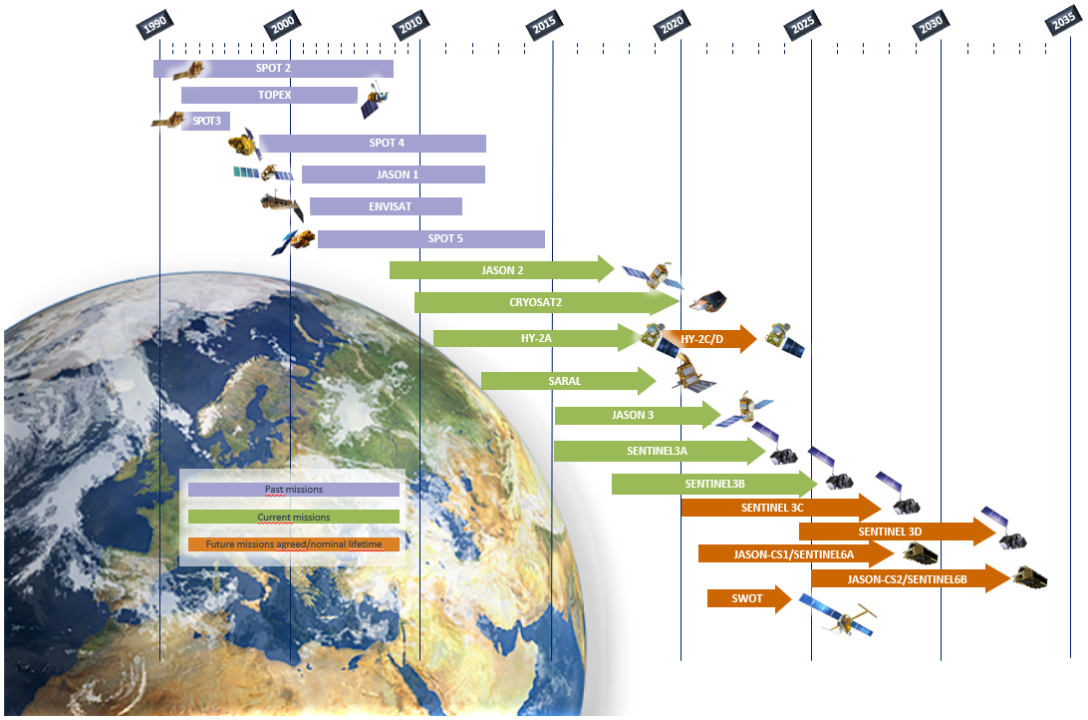

Figure 1.1 *Satellite altimetry timeline. Figure replicated from International-Altimetry-Team (2021) (updated to 2018)*

Sea Surface Temperature

Satellite Sea Surface Temperature (SST) is the other important remote sensed observation in physical oceanography. SST is strongly related to air-sea interaction processes and also characterizes the mesoscale variability of the upper ocean. A set of different satellites equipped with infrared or microwave sensors measure the sea temperature in the first few millimetres (”skin temperature”). Infrared radiometers onboard weather satellites can provide very high-resolution observations (up to 1 km), with high accuracy but affected by clouds. On the other side, microwave radiometers are not so affected by clouds and can measure SST in all weather conditions but with a lower spatial resolution (around 25 km). The difference in the measurements between both kinds of sensors can cause biases when comparing or blending products. This problem is addressed by processing ”Foundation SST”, which relates to the surface temperature without accounting for the diurnal heating (Donlon 2009). In the last decade, the Group of High-Resolution SST (GHRSST) Project (https://www.ghrsst.org) has blended different satellite products, producing a suite of high-resolution (< 10 km), level-4 (gridded) SST products that synthesize SST observations from various satellites, some also including in-situ SST observations (Lee and Gentemann, 2018).

Argo floats

While satellites provide an invaluable observation of the ocean surface, it is not enough to understand the entire 3D structure of the ocean. The Argo program is a major component of the Global Ocean Observing System (GOOS). Initiated in the late 90s, it provides since 2006 global in-situ measurements around the globe, representing the second revolution in operational oceanography (Le Traon, 2013). Argo floats measure temperature and salinity profiles (also biogeochemical observations lately) down to 2000 m. Floats have a nominal period of 10 days (5 in the Mediterranean Sea), meaning that they provide a high-resolution profile measurement with this frequency. The floats are designed to drift at a depth of 1000 m, gather data in up-cast with the designed frequency and send it to the distribution centres via satellite, providing near-real-time data. By November 2018, Argo had provided 2,000,000 profiles since the program began and a comparable number of velocity drift estimates at 1000 m depth (Roemmich et al., 2019). Nowadays, Argo profiles are used in combination with satellite altimetry and SST to constrain ocean models forecasting (Oke et al., 2009; Guinehut et al., 2004, 2012).

Paradigm change: Coastal Observing Systems

Satellite observations and Argo in-situ data together with moorings and tide-gauges conform the global observing system. Regional observations have been carried out since much before. Campaigns on board research vessels were the traditional way of studying the ocean. Researchers have traditionally designed experiments to study specific processes in a regional area. This traditional way has brought a broad scientific knowledge, and it is still useful; however, it is inevitably affected by several logistic problems. Campaigns are expensive, require many people, depend on weather circumstances and are limited in space and time.

In opposition to the traditional vessel experiment the tendency is to diversify, using new technologies and multi-observing platform systems, in what has been called the ”paradigm shift” (Delaney and Barga, 2009; Tintoré et al., 2013). This shift is twofold: on one side, the use of new technologies permits measuring the different scales in the ocean, capturing a wide variety of processes, from large scale to meso and sub-mesoscale, and obtaining from local to regional and global observations (e.g., Pascual et al. (2017)). At the same time, efforts are driven towards the collection and dissemination of data as quickly as possible. This clearly benefits the development of operational oceanography since near-real-time observations are available to run, validate and correct models. Presently, regional and global ocean observatories around the globe are dedicated to the development of multi platform coastal observing systems that combine different technologies and instrument for the observation of physical and biogeochemical variables of the ocean.

The main conventional components of the Global Ocean Observing System (GOOS) include satellite altimetry and sea-surface temperature (SST), Argo floats, eXpendable BathyThermographs (XBT), surface drifting buoys and moorings. These platforms partly resolve the mesoscale ocean circulation (Oke et al., 2015a). Regional ocean observing systems are being developed around the world under the auspices of the GOOS Regional Alliance [1], which is comprised of national and institutional efforts that co-

[1] www.iocgoos.org

operate regionally to complement the GOOS. Regional observation platforms include mooring arrays, land-based high-frequency radar (HFR) arrays and repeat glider deployments and are organised under projects such as EuroGOOS [2], USGOOS [3], IOOS [4], IOGOOS [5] and IMOS [6] (Oke et al., 2015b).

Gliders

Since the visionary "dream" of Stommel (1989), and following the community recommendations from the last decade (Testor et al., 2010), underwater autonomous vehicles (UAV) such as gliders, have become an essential element of regional ocean observatories. By controlling their buoyancy with an oil pump in its front part, gliders can navigate autonomously due to its hydrodynamics with a low power consumption. Given some navigation waypoints, gliders can travel for several months, doing up and down casts, gathering high-resolution observations until down to 1000 m. Once every certain dives, the platform sends the measured data correcting its trajectory if necessary, to track areas of interest and adapting to safety conditions (Garau et al., 2009). Gliders can measure a wide spectrum of processes being suitable to sample meso and submesoscale processes. They have been used to measure internal waves, turbulent dissipation and biogeochemical processes dominated by submesoscale (Rudnick, 2016). They can also be used to identify different water masses (Pascual et al., 2017; Heslop et al., 2012), and are capable of improving ocean model predictions via data assimilation (Melet et al., 2012; Shulman and Paduan, 2009; Pan et al., 2014). Gliders permit the performance of endurance lines and navigation in meteorological conditions and regions where vessels are unable to work, also reducing costs compared against traditional campaigns. Gliders also permit an adaptive sampling, changing their route along the mission, making it possible to collect data at optimal locations to maximize their information content, based on specific undergoing processes of special interest (Mourre and Alvarez, 2012; Lermusiaux, 2007).

High-frequency radars

High-frequency radars (HFR) are a shore-based remote sensing technology that measure ocean surface currents, wave fields, and winds near coastal areas. They are increasingly being installed in multiple coastal observatories all over the world due to their cost-effectiveness (Roarty et al., 2019; Rubio et al., 2017a). They allow real-time measurements providing a new, detailed, and quantitative description of physical processes at the marine surface (Mantovani et al., 2020).

HFR systems are generally composed of two or more antennas. Each antenna emits pulses at different ranges, measuring radial surface currents toward or away from the radar over many small sectors surrounding the radar. The sectors typically extend over a few degrees of azimuth and a few kilometres in range from the radar. The estimation of the velocity in each sector of the sea surface is done exploiting the Bragg resonance phenomenon (Crombie, 1971). It consists of measuring the spectrum of the backscattered signals over periods of several minutes, computing the corresponding Doppler

[2] www.eurogoos.org
[3] www.ioc-goos.org/usgoos
[4] www.ioos.noaa.gov
[5] www.incois.gov.in/Incois/iogoos/intro.jsp
[6] www.imos.org.au

frequency shift. The echo scattered from the sea is amplified when the transmitted radio signal encounters a surface wave that has a wavelength that is half the wavelength of the radio signal. The total surface current vectors at points on the sea surface are computed from radial measurements from two or more radars with overlapping coverage.

HFR complement satellite altimeter observations which are limited to larger scales and suffer limitations when approaching the coast (Vignudelli et al., 2019; Pascual et al., 2015). The capability of HFR to give realistic observations of surface currents has been widely validated (Chapman et al., 1997; Emery et al., 2004; Paduan et al., 2006). HFRs are important sources of data for understanding the coupled ocean-atmosphere system and the different coastal circulation processes like ocean waves, wind-induced currents, tidal flows, (sub)mesoscale variability, and inertial oscillations. They can be used to validate numerical models (Aguiar et al., 2020; Mourre et al., 2018; Lorente et al., 2021b), analyze Lagrangian dynamics (Hernández-Carrasco et al., 2018) or constrain numerical models via data assimilation (DA) (Vandenbulcke et al., 2017; Iermano et al., 2016a; Janeković et al., 2020a; Hernandez-Lasheras et al., 2021). Readers are referred to Paduan and Washburn (2013) for a more extended review.

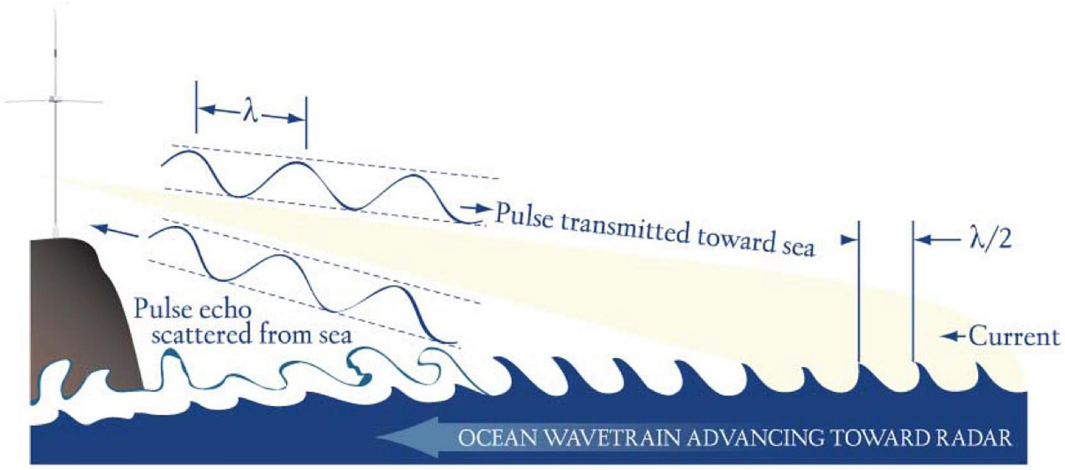

Figure 1.2 *HFR signal emission scheme depicting the Bragg scattering process that allows for ocean current measurements (from Roarty et al. (2019))*

1.3 Ocean Circulation Models

The increasing computational capabilities have boosted the development of numerical ocean models in the last decades similarly as the advancement in atmospheric prediction. Ocean circulation models are 3D representations of the movement of the oceans. They consist in solving a set of partial equations that describe the dynamics of the ocean, the so-called Navier Stokes equations, with some assumptions to make them solvable. Given initial and boundary conditions and some forcing fields (winds and radiation in the surface boundary layer) the ocean state evolution is estimated by solving the discretized diferential equations on a grid (Chassignet, 2021).

Ocean circulation modelling is a relatively modern field, with the first algorithmic formulation proposed in the 1960s by Bryan and Cox (1967) (a historical review can be found in McWilliams (1996)). Over the past decades, improved understanding, numerics, grid configurations, spatial discretization, parameterizations, data assimilation,

environmental monitoring, and process-level observations/modelling have led to significant advances in ocean circulation modelling (Fox-Kemper et al., 2019).

Models are used for diverse applications. They can be employed as an experimental tool for improving our mechanistic understanding of the ocean; for climate modelling when integrated into Earth System Models; for seasonal to decadal forecasts if coupled to atmospheric circulation models; and for producing short-range ocean forecasts and for ocean reanalysis when integrated in a data assimilation system (Le Sommer et al., 2018). This last activity is the most related to the operational oceanography framework of this Thesis.

Ocean circulation modelling can also go from global to regional, basin-scale or coastal applications. Actual global models can resolve mesoscale features, with resolutions of the order of 1/12 degree (9.25 km in the equator and around 4.5 km in subpolar latitudes), as the actual Copernicus Global reanalysis (Lellouche et al., 2021). Actual regional models can have resolutions of the order of a hundred meters, resolving submesoscale processes in coastal areas (Sotillo et al., 2021; Capet et al., 2020). They are nested into the global models, which provide the initial and boundary conditions. However, caution is needed when downscaling, as important issues may arise from the fact that the large-scale solution is unbalanced with respect to the local physics of the embedded model due to the different resolutions, bathymetries, numerical boundary conditions, etc (Kourafalou et al., 2015b).

The nature of the ocean motion is chaotic, and there are many non-linear processes unresolved by the models. Processes that are not resolved due to the scale of the grid are parametrized, usually using analytical formulations. Even if the initial ocean state was perfectly known, predictions would be inevitably affected by errors from multiple sources, as boundary conditions, forcing fields, bathymetry smoothness, inaccurate parametrizations or unresolved scale processes, affecting the integration of the model solution and making the forecast diverge from the real ocean state. This is why in operational oceanography, it is extremely important to correct the forecasts using observations (Wilkin and Hunter, 2013; Onken et al., 2008; De Mey-Frémaux et al., 2019).

1.4 Data Assimilation

Data assimilation (DA) aims at merging dynamical models with observations in an optimal way to provide reliable ocean state estimates. Ocean data assimilation is increasingly recognized as crucial for the accuracy of real-time ocean prediction systems and historical re-analyses (Moore et al., 2019). In general, DA exploits the models as spatio-temporal interpolators of the data, and the data guide the models toward the true evolution of the system. Effective operational oceanography relies on the ability to assimilate massive amounts of data gathered by monitoring systems in real-time into general circulation models (Hoteit et al., 2018).

Observations are obviously not exempt from errors, apart from the obvious instrumental ones. Many times we have to deal with synopticity issues, especially in 3D DA methods, when we consider all observations as synchronous. Using spatially or temporally averaged variables is also typical, mainly when using mapped observation, as SST L4 level products. Sometimes in-situ profiles are considered representative of an area in which small scale structures take place, or on the other side, a super-obing process is applied, generating a single observation from multiple ones, due to computational limitations. All the errors that we might be assuming by making these considerations

are enclosed in what we call representation errors. In general, the representation error can be understood as the component of observation error due to unresolved scales or attributed to any physical processes appearing in the observations but not in the model (Oke and Sakov, 2008). We use it when we don't want the analysis to fit so closely the observation.

The solution of the DA analysis will be an ocean state that should lie between the a prior state (or background) and the observations. This is generally conceptualized in terms of Bayes' theorem, the axioms that govern probability (Bayes, 1763), as follows: Given a prior information about the laws governing the ocean state in the form of a model, an a prior state estimate from that model, and direct, but incomplete, ocean observations, an a posterior state estimate is computed that weights all available information according to the hypothesized uncertainties in the model and observations Moore et al. (2019). In a Bayesian framework, the optimal state estimate is that which coincides with the maximum a posterior probability (Wikle and Berliner, 2007). To make this approximation some assumptions should be made. In particular, model and observation errors are random, unbiased and follow a Gaussian distribution.

Usually, the ocean state is regularly updated every certain number of days. DA methods are mainly divided into variational and sequential methods. Both types of methods are used indistinctly, with satisfactory results, in different assimilation systems around the world (Martin et al., 2015).

In variational methods the ocean state is corrected by minimizing a given cost function that measures the model to data misfit (Bertino et al., 2003). Variational methods can be divided in: i) 3DVar, if the analysis should be adjusted to observations in a single instant, considering them synoptic or ii) 4DVar, if the model trajectory is the solution that best fits the observations, according to its own dynamics while minimizing the cost function. 4DVar provides theoretically a better solution but is much more computationally expensive, as it requires the execution of an adjoint and tangent linear models with multiple loops.

Among the sequential methods, the most popular used in oceanography are ensemble based methods rooted in Kalman Filter (KF) theory. In these methods an ensemble of simulations is used to calculate the model background covariances. While EnKF considers an ensemble that varies along the analysis time, EnOI supposes a simplified version where the ensemble members are static for each analysis, having a background covariance that does not change over time. The EnOI can be an attractive approach to save computer time. Once the stationary ensemble is created, only one single model integration is required in addition to the analysis step where the final update cost is reduced. The method is numerically extremely efficient but it will always provide a suboptimal solution compared to the EnKF (Sakov and Sandery, 2015). However, when using a bigger number of ensembles and for cases where localization is required, EnOI can be a good alternative and even outperform EnKF (Oke et al., 2007).

Sequential assimilation methods are affected by initialization issues when restarting the model after the analysis step, as primitive equation models are sensitive to discontinuous changes in their model fields (Oke et al., 2002). These discontinuities may introduce artificial waves or structures in the model that affect the quality of predictions. Different strategies have been proposed to address this problem (Sandery et al., 2011; Yan et al., 2014). In this Thesis, we will also deal with this problem, trying to propose different strategies to restart the model after performing an analysis.

Notice that hybrid methods and data-driven approaches, also combining neural-

network techniques with classical dynamical model procedures are increasingly being explored in the last years. A complete review of DA methods and future perspectives can be fund in Carrassi et al. (2018).

1.5 Model-Observation synergies through observation impact studies using Data Assimilation

Global programs developed in the last 20 years intend to foster cooperation between different countries and institutions. The problems we try to address are global, and the difficulties in sampling the ocean make it even more necessary to strengthen this cooperation. In Europe, several programs have been developed in the last year trying to build a sustained network of coordinated ocean observing systems (Farcy et al., 2019a; Fujii et al., 2019).

Regional ocean modelling relevance has also been highlighted in numerous studies. Ocean circulation models corrected with real-time observations through DA are of extreme importance to answer scientific and societal needs. For instance, search and rescue operations, marine traffic or port services highly rely on proper operational forecasts.

However, some efforts are still needed to improve the model forecasts and the ocean observing system in general. The development of data assimilation systems able to incorporate observations such as gliders and HFR is crucial to constrain submesoscale structures and be able to understand coastal processes and respond to operational oceanography necessities.

The OceanPredict (formerly GODAE OceanView; Davidson et al. (2019)) network is an international research and development initiative to develop and increase the impact of ocean prediction. However, despite the global programs, a common criticism within the community is the lack of cooperation between the observational and the modelling communities. There may be a lack of combined work in terms of basic science, decision making, or infrastructure planning. A call to join efforts and strengthen synergies between different communities is encouraged (De Mey-Frémaux et al., 2019). Data assimilation, with Observing system experiments (OSE) to assess model-observing system capability and Observing System Simulation Experiments (OSSEs) to evaluate future ones, provides a very valuable framework for it.

Observing System Experiments

OSEs are experiments designed to assess the impact of real observations in a forecast model. Normally, the impact of a particular observing source or platform is quantified with data-denial experiments. I.e., a series of simulations are run, with and without the observations to evaluate and the results are compared to a free-run simulation with no data assimilation. OSEs impact studies can help to the efficient design and optimization of observing systems for scientific and operational uses (Fujii et al., 2019). Multiple examples of this kind of experiments can be found in the literature, however the results are inevitably system-dependent Schiller et al. (2015). For instance, (Oke et al., 2015a) provides an overview of the impact of observations on global ocean forecasts and reanalysis. Generally, all global models assimilate observations from satellite, typically SLA and SST, which are given with full coverage and constrain the general circulation and main features. As mentioned before, Argo floats also provide near-real-time

global profile observations, helping to constrain the density fields in depth. Besides, the impact of observations from XBT lines, marine mammals, sea ice-concentration (Xie et al., 2016) or moorings as the PIRATA array (Bourlès et al., 2008) have also been analyzed in different global systems (Lellouche et al., 2013). OSEs of global systems, include different models, data assimilation schemes and a wide variety of data (e.g., Lellouche et al. (2018); Chassignet et al. (2007)). A review of the main OceanPredict systems can be found in Martin et al. (2015).

Regional basin-scale OSE have also been performed following the goals of OceanPredict as Oke et al. (2008), were a long-run reanalysis simulation of the Bluelink system, in Australia, is analyzed. Tanajura et al. (2020), presents a series of OSEs in the RODAS system, which covers the Brazilian coast and most part of the south Atlantic ocean, using different sets of observations. They evaluate, for instance the effect of including or not Argo floats, and the impact of sea-surface-salinity (SSS) or altimetry in their system.

In a more regional context, OSEs have been performed to evaluate how high-resolution models and observations are able to better reproduce the finer-scale processes. For instance, Kerry et al. (2016, 2018) performed a reanalysis simulation in the eastern Australian coast, using all the available observation from the IMOOS system, such as gliders, HFR, XBT, moorings... With their DA system they were able to evaluate the impact of each source of data in their simulation. A similar approach is followed in Wilkin and Hunter (2013), were the MARACOOS observing system is assessed, using different model configurations.

In coastal regions OSEs and OSSEs are very helpful to study, reconstruct and characterize coastal ocean processes and its variability Kourafalou et al. (2015a). In this regions where different scales interact, the integration of different kind of observing sources into high resolution models is crucial to respond to operational oceanography requirements. Stanev et al. (2015) provides a good example of how the integration of multiple observing sources from the COSYNA observing network into a coastal modelling system to correct tides and circulation processes occurring at a local scale.

Several OSE studies have been done in the Mediterranean Sea using multiple different observing sources. The impact of gliders was evaluated by Mourre and Alvarez (2012) in a coastal model in the Ligurian Sea, showing the benefits of sampling the are with an adaptive-sampling-driven glider compared one with a pre-defined route. More examples of glider OSEs in the Mediterranean sea can be found in Dobricic et al. (2010) or in Hayes et al. (2019), where the impact of assimilating glider temperature and salinity profiles was assessed in a six month long simulation. Other observation impact studies have been performed to evaluate the use of Ferry-Box (Korres et al., 2014), for instance, or from different types of SST products (Storto and Oddo, 2019).

OSSEs

Conceived as an expansion of OSEs, OSSEs are very useful approach for the design of an observing network, its expansion or improvement and to assess different future scenarios. OSSEs are ocean observation impact studies in which we use a realistic model simulation and consider it as the true state of the ocean system we are studying. In this approach, we can generate pseudo-observations from this realistic simulation, often called Nature Run (NR), so as to emulate the observing system to study. OSSEs approaches have been developed and widely used in atmospheric science

and had only recently been adopted for the ocean. Halliwell et al. (2014) is the first study that performed a rigorous design criteria for oceanic OSSEs, analog to the design experiments established by Atlas (1997) for the atmospheric systems. This rigorous validation implies that the NR should be validated to give realistic simulations and the pseudo-observations generated should resemble the real ones. Furthermore, it should be compared to a real OSE, used as reference to assess that the impact in both frameworks is consistent. Although some OSSEs have been published before, most of them do not follow full-fledged data assimilation system procedure, but most likely some field reconstruction technique, as interpolations, Kalman filters, or empirical orthogonal functions (EOF) based methodologies Guinehut et al. (2004); Ballabrera-Poy et al. (2007); Sakov and Oke (2008).

More recently, there have been a few studies following the criteria established by Halliwell et al. (2014) as in Halliwell et al. (2020) where an underwater glider array is assessed to improve ocean model initialization for tropical cyclone forecasting. Using a similar approach Gasparin et al. (2019) studied how deeper observations of Argo floats will impact a global ocean reanalysis, and Benkiran et al. (2021) assessed the impact of the assimilation of data from the future SWOT satellite mission in a global-high-resolution model. In a regional context, Alvarez and Mourre (2012) performed several OSSEs to determine an optimal glider and mooring sampling strategy in the Ligurian Sea. Other OSSEs in the Mediterranean Sea include the assessment of a Fishery Observing System (FOS) network (Aydoğdu et al., 2016) or the impact of satellite and in-situ observations in Underwater Acoustic Predictions (Storto et al., 2020). However there is still a lack of this kind of experiments evaluating the impact of high-resolution observations in regional and coastal models.

1.6 The Mediterranean Sea

The Mediterranean Sea is a semi-enclosed sea, in the mid-latitudes of the northern hemisphere (30ºN - 46ºN), which spans meridionaly from 6ºW to 36ºE. It is connected to the Atlantic ocean through the strait of Gibraltar, a narrow passage that separates the Iberian peninsula from the African continent (Fig. 1.3). It is about 13 km in its most wider part, with a minimum 300 m depth section. The Mediterranean is also connected to the Black Sea, which is considered as independent, through the Strait of Dardanelles and the Marmara Sea, in its north-eastern part (around 28ºE). The Mediterranean Sea has a mean depth about 1500 m and a deepest part situated in the Ionian Sea, east of Sicily island, with more than 5000m. It is divided in two basins, the Eastern Mediterranean and the Western Mediterranean (WMED) through the strait of Sicily (~ 400 m depth), which are at the same time divided into several sub-basins (Millot and Taupier-Letage, 2005).

From a scientific point of view, the Mediterranean Sea is interesting due to the many fundamental processes that take place in the ocean and occur here at a smaller scale. For instance, deep water formation processes, thermohaline circulation, different water mass interactions, baroclinic instabilities or mesoscale activity (Malanotte-Rizzoli et al., 2014). Its latitude and bathymetry, makes that some of these processes happen in a smaller spatial scale, which make them easier to sample and study. This is why among oceanographers it is sometimes referred as a miniature ocean laboratory (Bethoux et al., 1999). The Mediterranean Sea is a concentration basin, where the loss due to evaporation exceeds the freshwater inputs from precipitation and river run-off. Circulation in

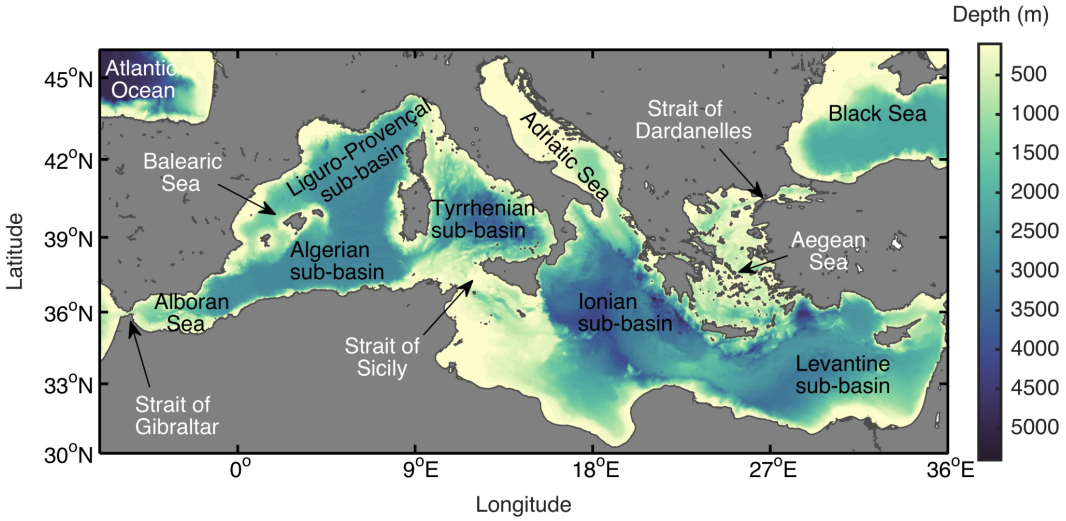

Figure 1.3 *Mediterranean Sea bathymetry map with its different sub-basins and main straits indicated. (borrowed from Piñeiro (2021))*

the whole Sea, and also in each particular basin, is anti-clockwise, and the main water exchange occurs across the strait of Gibraltar, were two flows with very different thermohaline characteristics are superposed: A fresher (less saltier) water from the Atlantic ocean on top and a saltier water in depth.

Moreover, the Mediterranean Sea is impacted by a big antrophogenic pressure. More than 100 million people leave along its coasts (CIESIN information, at 2021), is a transit of many of the most important maritime routes and receives the runoff of many big rivers (e.g., Nile, Ebro, Rhone, Po), which are sources of intense pollution (Cózar et al., 2015). It is also the main touristic area in the world, source of intense fishing activity, apart of other numerous economical interests. Many ecological processes of interest also take place in the Mediterranean Sea. Besides, it is hot geopolitical area, where many migrant routes happen, and more than 5000 people loss their lives every year (Steinhilper and Gruijters, 2017).

1.7 SOCIB. The Balearic Islands Coastal Observing and Forecasting system

SOCIB is a Coastal Ocean Observing and Forecasting System located in the Balearic Islands (Tintoré et al., 2013). SOCIB works operationally since 2012, gathering, treating and disseminating ocean observations and forecasts in the Western Mediterranean Sea. It is aligned with the marine strategy by the European Commission mandate and tries to respond to societal and scientific requirements. SOCIB is funded by the Spanish Science Ministry, the Balearic Islands Government and the Spanish Research Council (CSIC) and works in partnership with other research institutions in multiple European and international programs and projects, as CMEMS, JERICO or EuroSea, in line with the some of the objectives of the work developed in this Thesis.

The activities developed in SOCIB are divided in three main branches, closely related one with each other. First, the observing system, composed by different facilities, which include a research vessel, gliders, HFR, Lagrangian platforms, moorings,

weather stations or tide gauges. Secondly, the data center, which is in charge of the data curation and dissemination. Finally, the modeling and forecasting facility (MFF), which provides wave, ocean circulation and meteotsunamis forecasts daily in an operational context, integrating as much as possible the observational information provided by the other facilities for the validation and assimilation in the operational models. The MFF also provides scientific and technical advice for different stakeholders and performs scientific research, as the work presented here, with the aim to advance the knowledge and predictive capabilities, according to SOCIB strategic plan.

1.8 WMOP: The Western Mediterranean Operational System

1.8.1 Model

The model that is used all along this Thesis is the Western Mediterranean OPerational system (WMOP, Juza et al. (2016); Mourre et al. (2018)) developed at SOCIB. It is a high-resolution regional configuration of the ROMS (Regional Ocean Modelling System) model (Shchepetkin and McWilliams, 2005) for the western Mediterranean Sea. ROMS is a three-dimensional free-surface, sigma coordinate, primitive equations model using split-explicit time stepping with Boussinesq and hydrostatic approximations. The spatial coverage of WMOP spans from Gibraltar strait on the West to the Sardinia Channel on the East (6ºW-9ºE, 35ºN-44.5ºN) with a horizontal resolution around 2 km and 32 vertical sigma levels, that stretch from the surface to the bottom boundaries.

Different model configurations have been developed in the past to optimally setup the system, testing initial and boundary conditions, and different parameters. The simulation used here is initialized from and nested within the larger scale Copernicus Forecasting System (CMEMS MED-MFC), with a 1/16º horizontal resolution (Simoncelli S., 2017). The vertical mixing coefficients are set using the Generic Length Scale (GLS) turbulence closure scheme (Umlauf and Burchard (2003), with parameters p=2.0; m=1.0; n=-0.67 as in line 1 of their Table 7). The bathymetry is derived from a 1' database (Smith and Sandwell, 1997). The Spanish Meteorological Agency (AEMET) provides HIRLAM model (Undén et al., 2002) atmospheric outputs every 3 hours at $1/20^{\circ}$ horizontal resolution, that are used to force the ocean model. These fields are used to compute surface turbulent and momentum fluxes through bulk formulae. Atmospheric pressure forcing is neglected to avoid SSH high-frequency variability issues. Inflows from the six major rivers in the region are considered as point sources, using daily climatological values. Tides are not considered in the model.

The free-run hindcast, spanning the period 2009 - 2019, has been deeply evaluated using satellite and in-situ observations (Mourre et al., 2018; Aguiar et al., 2020). The mean surface height and circulation over the year 2014, derived from altimetry and from the free run simulation is illustrated in Figure1.4. In particular, it is found to properly represent the mean surface geostrophic circulation over the basin, and the main features which are the Alboran Gyres, the Algerian Current along the African coast and its associated instabilities, the Northern Current along the French and Spanish Coast and the Balearic Current flowing north-eastwards north of the Balearic Islands. This is in agreement with both the historically established regional surface circulation (Millot, 1999) and the more recent average estimates provided by the Mean Dynamic Topography (Rio et al., 2014).

The outcomes of this Thesis are also used to develop the WMOP operational system,

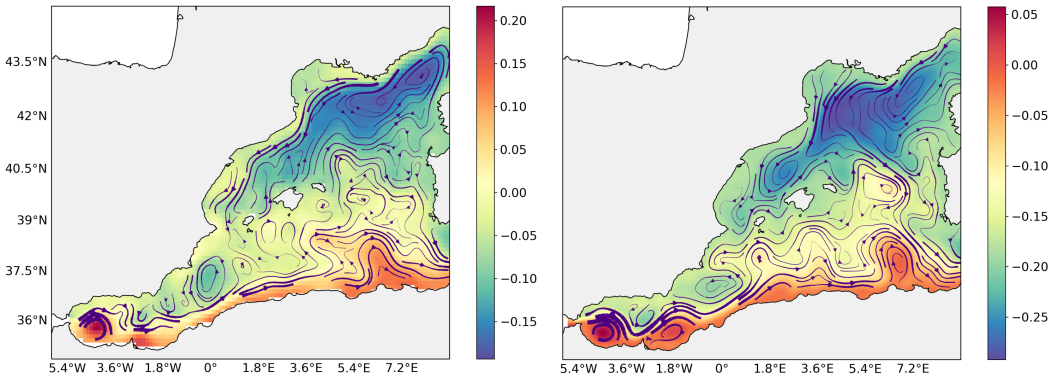

Figure 1.4 *Sea surface height and geostrophic current streamplot derived from altimetry (left) and the WMOP free-run hindcast simulation (right) for the year 2014.*

that is used to produce daily forecasts of the regional ocean circulation. The model is used for a wide range of applications including search-and-rescue and analysis of plastic, parasite or larval dispersion (Calò et al., 2018; Ruiz-Orejón et al., 2019; Cabanellas-Reboredo et al., 2019; Compa et al., 2020; Torrado et al., 2021; Kersting et al., 2020; Révelard et al., 2021).

1.8.2 Data Assimilation system

A Data Assimilation scheme has been developed for the WMOP system, in order to to ingest data from multiple types of observing sources, including large scale satellite observations all over the domain or high-density data in a limited area. The scheme is a form of the EnOI, which is a 3D sequential assimilation method widely used in operational systems. Theoretically, EnOI is suboptimal, compared with more complex methods as the Ensemble Kalman Filter or the 4Dvar (Evensen, 2003; Counillon and Bertino, 2009). However, it is a cost-effective alternative, since permits the use of a large ensemble size together with localization, being able to even outperform EnKF (Oke et al., 2007).

In ensemble methods, an ensemble of model realizations is used to compute the background error covariances (BECs), that in the case of EnOI are stationary for each analysis. The estimation of the ensemble is a critical point, since the value of the residuals (innovation after the analysis) directly depends on the BECs and in the observation error, as explained below. Instead of generating an ensemble of simulations for each analysis, in our approach we estimate the BECs by sampling three long-run simulations of the WMOP with different initial and boundary forcing and varying momentum and diffusion parameters. Each ensemble member is randomly extracted from the three different long-run simulations within a temporal window of 90 days centered on the day of the analysis for the different years covered by the three long-run simulations. The seasonal cycle is removed from the multivariate fields before computing the ensemble anomalies to limit the effects of large scale correlations, mainly in terms of surface temperature. This way, we obtain multivariate, inhomogeneous and anisotropic 3-dimensional model BECs characteristic of the mesoscale variability. A domain localization avoiding spurious long distance correlations effects is used. This way, an

independent analysis is performed for each water column of the model domain, considering only the observations within the localization radius. These are the reasons why the system is referred to as local multi-model Ensemble Optimal Interpolation (EnOI).

The WMOP-DA system consists of a sequence of analyses (model updates given a set of observations) and model forward simulations. For each analysis, the state vector $\mathbf{x} = (\mathbf{T}_{i,j,k}, \mathbf{S}_{i,j,k}, \mathbf{u}_{i,j,k}, \mathbf{v}_{i,j,k}, \mathbf{SSH}_{i,j})^{\mathbf{T}}$, contains the model trajectory, i.e., the prognostic model variables at all wet gridpoints i, j, k. During the analysis step, the state vector $\mathbf{x}^a$ is updated according to Eq. (1.1), where $\mathbf{x}^f$ is the background model state vector, $\mathbf{H}$ is the linear observation operator projecting the model state onto the observation space and $\tilde{\mathbf{K}}$ is the Kalman gain estimated from the sample covariances (Eq. 1.2). $\mathbf{y}$ is the vector of observations. Matrices $\tilde{\mathbf{P}}^f$ and $\mathbf{R}$ are the error covariance matrices of the model and the observations, respectively. $\tilde{\mathbf{P}}^f$ contains the background error covariances (BECs).

$$\mathbf{x}^a = \mathbf{x}^f + \tilde{\mathrm{K}}(\mathbf{y} - \mathbf{H}\mathbf{x}^f), \tag{1.1}$$

$$\tilde{\mathbf{K}} = \tilde{\mathbf{P}}^f \mathbf{H}^T (\mathbf{H}\tilde{\mathbf{P}}^f \mathbf{H}^T + \mathbf{R})^{-1}, \tag{1.2}$$

1.9 Structure and Objectives

Within this context, the main objective of this Thesis is to evaluate the impact of new coastal observing datasets on the prediction of the high-resolution WMOP model in the Western Mediterranean Sea. Specific aspects are investigated in the different chapters. The chapters presented here are self-contained, as they are written in the form of articles already published (Chapters 2 and 3) or in preparation for a future submission (Chapter 4). Thus, some of the information may be repeated along the document.

The objective in Chapter 2 is to evaluate the trade-off of assimilating glider temperature and salinity observations in opposition to traditional sampling campaigns on-board of a research vessel (RV). Special focus is put on comparing the impact of a fleet of gliders sampling a reduced coastal area compared to a traditional CTD sampling from a RV. The chapter tries to deepen the importance of measuring different scale processes in the ocean and the possible advantages that glider data may bring. For it, we harness the observations gathered in a massive campaign west off Sardinia coast, the REP14 experiment. We evaluate the effects of assimilating traditional observing sources into a regional model, as satellite SLA and SST, and Argo profiles, all over the domain with dense high-resolution campaign observations.

In Chapter 3, we assess the capacity to correct surface currents assimilating HFR observations in the Ibiza Channel (Western Mediterranean), comparing different types of HFR observations (radials and totals) and initialization methods after analysis. The objective here is to evaluate the importance of this high-resolution surface current observation to constrain the circulation in a coastal model. The Lagrangian validation performed measures the potential impact of HFR DA to predict the trajectory of drifting buoys. Correction of small scale structures in the area is relevant for search and rescue operations in the area and understanding of ecological processes, among others.

As a step further, an HFR OSSE experiment is performed in Chapter 4. Using a calibrated OSSE framework, the impact of a couple of new antennas extending the actual coverage is evaluated in the Ibiza Channel. The OSSE framework is first validated by comparison with the OSE performed in the previous chapter. As a second objective, we evaluate the impact of DA on the transport in the Ibiza Channel. A Lagrangian

validation is proposed based on the computation of Lagrangian Coherent Structures (LCSs) regulating the region's transport. We also try to answer the question of whether model fields corrected with DA are valid to perform this kind of analysis and useful to reconstruct the LCSs present in the observed ocean state.

Finally, Chapter5 summarizes the main findings of the Thesis and concludes it proposing new research lines that should be pursued to further improve the predictive capability of numerical models exploiting the present potential of the Ocean Observing Systems.

Chapter 2

Dense CTD survey versus glider fleet sampling: comparing data assimilation performance in a regional ocean model west of Sardinia

Abstract

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

2.1 Introduction

Short-term regional ocean prediction is important to respond to maritime emergencies related to search-and-rescue or accidental contamination, for maritime security or

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

In order to constrain errors and remain as close as possible to reality, models must be feeded with different kinds of observations. Satellites play a key role, providing regular near real time data of surface variables such as temperature and sea surface height. Water column measurements are more scarce. The Argo program provides routine temperature and salinity profiles at regular intervals, but the distance between floats is insufficient to monitor the mesoscale and finer scale variability (Sánchez-Román et al., 2017). Dedicated campaigns providing underwater measurements from ship data or glider measurements provide complementary data over specific areas. Efficient DA systems should be able to advantageously combine large scale observations over a large domain with more dense, high resolution observations in specific areas. Traditionally, campaigns onboard research-vessels have been carried out to collect dense CTD data to initialize regional ocean prediction systems. However, campaigns are not always possible. They depend on ship availability, weather, access to the area of interest and they remain very expensive. Recent evolutions in technology allow to deploy autonomous underwater vehicles (AUV) such as gliders to collect dense hydrographic data over specific areas of interest (Testor et al., 2010; Ruiz et al., 2012; Rudnick, 2016; Liblik et al., 2016). Gliders are able to operate under hard maritime situations and to reach difficult access areas, with an overall cost reduced compared to traditional ship campaigns. Glider missions are typically planned to reach a series of locations commonly called waypoints, in order to track areas of interest and adapt to safety conditions (Garau et al., 2009). Their controllability also permits adaptive sampling procedures, changing their route along the mission with the objective to collect data at optimal locations to maximize their information content (e.g. Lermusiaux (2007); Mourre and Alvarez (2012)).

The potential of gliders to sample fine scale processes and to identify different water masses has been demonstrated (Pascual et al., 2017; Heslop et al., 2012), as well as their capability to improve ocean model predictions via DA (e.g. Melet et al. (2012); Shulman et al. (2008); Pan et al. (2014); Mourre and Chiggiato (2014)). The question arises whether the sampling offered by a fleet of several gliders is as useful as a traditional ship-based CTD survey for regional ocean forecasting applications. This is the question we are addressing in this chapter.

The rich dataset collected during the REP14-MED campaign is used for this purpose. REP14-MED took place in June 2014 offshore the Western coast of Sardinia (Onken et al., 2018; Knoll et al., 2017). Two research vessels tracked in parallel a 100 x 100 km area during a 20-day period, providing dense CTD sampling with a 10-km separation and continuous towed CTD measurements for limited periods of time. In addition, a fleet of 11 gliders were deployed performing back-and forth sections perpendicular to the coast with a 10km vehicle intertrack distance.

The REP14-MED experiment took place in the Sardinian sea, a region of the so-called Algero-Provençal basin of the Western Mediterranean Sea, one of the most dy-

namic areas of the entire Mediterranean sea (Olita et al., 2011; Millot, 1999). In this region, the surface layer is characterized by a water mass of Atlantic origin and a strong mesoscale activity. The region is strongly influenced by instabilities of the Algerian current, which generate intense anticyclonic eddies which can propagate northward towards the western Sardinian coast (Robinson et al., 2001; Testor et al., 2005; Escudier et al., 2016a). Such eddies can last from weeks to months. They are responsible for an intense mesoscale activity in the study region (Santinelli et al., 2008). A southward current flowing along the southern part of the Sardinian coast has also been evidenced in long term numerical studies (Olita et al., 2013) and in field campaigns contributing to episodically wind-induced advection of coastal water (Ribotti et al., 2004). At depth, eddies are also generated from the interaction between the Algerian Gyre and inflows of Levantine Intermediate Water (LIW) and Tyrrhenian Deep Water (TDW) coming from the Sardinia Channel (Testor et al., 2005).

The assimilation system used in this work follows an EnOI (Ensemble Optimal Interpolation) scheme (Evensen, 2003). This method provides a cost-effective approach when compared with more advanced methods such as EnKF or 4Dvar (Oke et al., 2008), which is suitable for operational implementations in regional ocean models. EnOI is a 3-dimensional sequential DA method. A stationary ensemble of model simulations is used to calculate background covariances. Contrary to the EnKF which requires to evolve an ensemble of simulations, a single model integration is only necessary between two analysis steps in the EnOI, making the method numerically efficient. The EnOI provides a suboptimal solution compared to the EnKF (Sakov and Sandery, 2015). However, it represents a good alternative allowing to use a large ensemble size together with localization when necessary (Oke et al., 2007). In this work, a Local Multimodel EnOI scheme is implemented. "Multimodel" represents the fact that the library of ocean states is built using different long-term model simulations. "Local" means that the EnOI analysis comprises some domain localization to reduce the impact of potential significant covariances associated with remote observations.

The chapter is organized as follows: Section 2.2 presents the observing and modelling frameworks, as well as the specific forecast experiment. Section 2.3 details the results, which are further discussed in Section 2.4. Finally, Section 2.5 concludes the chapter.

2.2 Data and Methods

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

2.2.1 REP14-MED experiment

The REP14-MED sea trial (Onken et al., 2018) was conducted in the framework of the EKOE (*Environmental Knowledge and Operational Effectiveness*) research program of the Centre for Maritime Research and Experimentation (CMRE, Science and Technology Organization - NATO). It is part of a series of sea trials dedicated to *Rapid Environmental Assessment* (REA), denoted by the acronym REP (*Recognized Environmental Picture*). Leaded by CMRE and supported by 20 partners, the trial took place during 20 days in June 2014, with NATO Alliance and the German Planet research ves-

sels (RV) conducting a joint survey over an aproximately 100×100 km area off the west coast of Sardinia (Figure 2.1). A massive amount of data was collected during the campaign with various oceanographic instruments, including CTD stations, towed Scanfish and CTD chain, ship mounted ADCP, shallow and deep underwater gliders, moorings, surface drifters and profiling floats. The sampling was divided into three legs. The time distribution of the collection of observations used in the present work is illustrated in Figure 2.2.

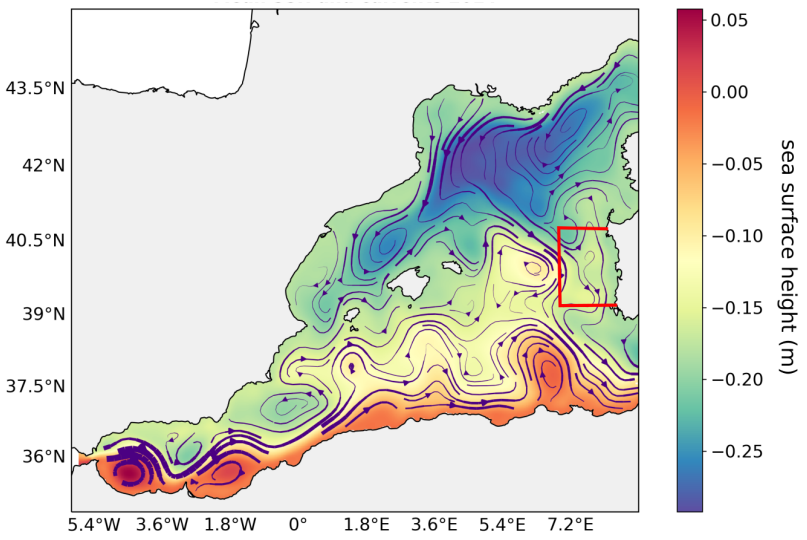

Figure 2.1 *WMOP free run model annual mean sea surface height and corresponding surface geostrophic currents for year 2014. The REP14-MED sea trial area is highlighted in red.*

During Leg 1, both RV conducted a parallel sampling of the target area, collecting CTD data with a 10km distance between stations over a 5 day period. During Leg 3, CTD data were collected with the same density, yet over a reduced spatial extension, providing very valuable data to validate the forecast experiments. CTD casts reached down to 1000m deep when possible. A few profiles even get deeper in order to characterize deep water masses. Additional towed Scanfish measurements of temperature and salinity down to 200m depth allowed to complete the characterization of the area during Leg 3. At the same time, and during the whole duration of the campaign, eight gliders were considered, travelling continously along back and forth transects perpendicularly to the Sardinian coast. Five of these gliders were deep gliders submerging to depths down to 800m, the remaning three were shallow water platforms collecting data in the upper 200m only. Each of these single transects was completed in about three days for each way. Notice that three additional gliders were deployed during the sea trial, but due to technical problems, duplication of the track and lack of processed data, they were discarded here. All glider tracks are approximately parallel to each other, with an intertrack distance around 10km, thus covering the target area. Figure 2.2 shows the position of CTD, glider and Scanfish data during Legs 1 and 3 of the sea trial.

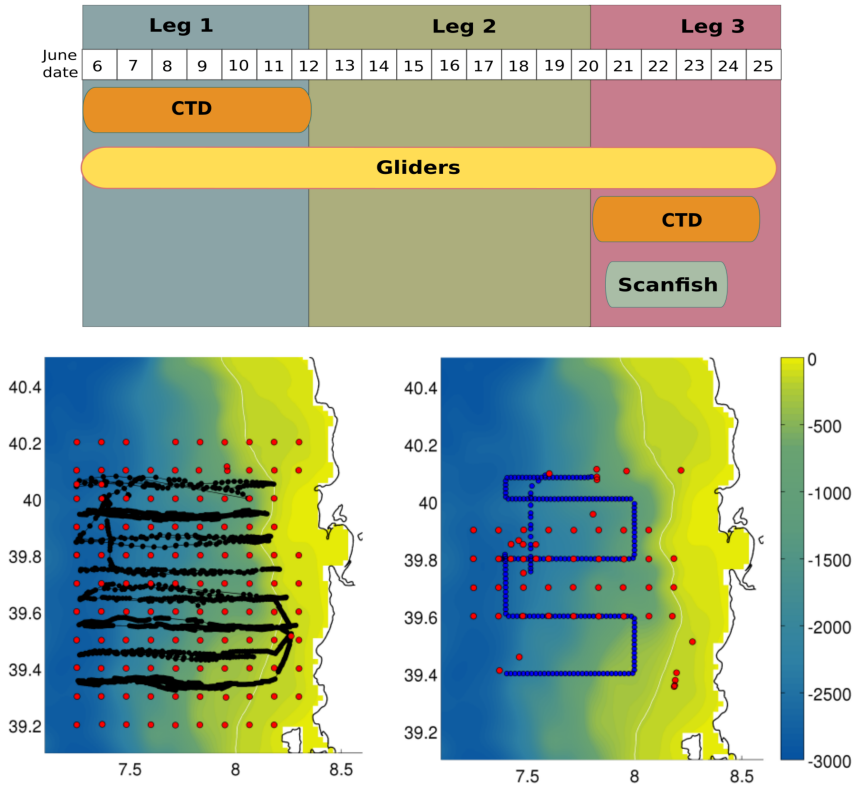

Figure 2.2 *Top panel: Sampling schedule in June 2014 of the REP14-MED sea-trial data employed in the present work. The spatial distribution of observations is illustraed in the two bottom panels. CTDs from Leg 1 (red) and gliders (black) observations used for assimilation are shown in the bottom left panel. Independent CTDs (red) and Scanfish (blue) gathered during Leg 3 and used for the validation are shown in the bottom right pannel. The colorbar indicates depth (m) and the white contour indicates the 200m isobath.*

2.2.2 Regional model configuration

The model used in this work is the Western Mediterranean OPerational model (WMOP, Juza et al. (2016)), covering a domain extending from Gibraltar strait to the Sardinia channel. WMOP is based on ROMS (Shchepetkin and McWilliams, 2005), a three-dimensional free-surface, sigma coordinate, primitive equations model using split-explicit time stepping with Boussinesq and hydrostatic approximations. WMOP is set-up with 32 vertical levels and a 2km spatial resolution. It is forced at the surface using the 3-hourly and 5-km resolution HIRLAM atmospheric model provided by the Spanish meteorological agency (AEMET). In this work, the initial state of the forecast experiments is provided by the simulation fields on 1 June 2014 of a seven-year long free run WMOP simulation spanning the period 2009-2015. This simulation uses initial state and boundary conditions from the CMEMS-MED reanalysis (Simoncelli S., 2017). In addition, several WMOP free run hindcast simulations were generated including modifications of the parent model used as initial and boundary conditions and some model parameters. These different simulations provide the library of ocean states used by the data assimilation system.

The model has been evaluated using satellite and in-situ observations (Mourre et al.

(2018)). The mean circulation of the free run simulation over the year 2014 is illustrated in Figure 2.1). In particular, it is found to properly represent the mean surface geostrophic circulation over the basin, in particular the main features which are the Alboran Gyres, the Algerian Current along the African coast and its associated instabilities, the Northern Current along the French and Spanish Coast and the Balearic Current flowing north-eastwards north of the Balearic Islands. Close to Sardinia, the mean circulation in the model is characterized by a south-eastward flow centered around 40N, which separates into two branches flowing northward and southward, respectively, when approaching the Sardinian coast, also giving rise to small eddies in the REP14-MED area. This is in agreement with both the historically established regional surface circulation (Millot, 1999) and the more recent average estimates provided by the Mean Dynamic Topography (Rio et al., 2014).

Moreover, we illustrate here the Sea Surface Temperature (SST) maps derived from the model and satellite data at the beginning of the REP14 period (Figure 2.3). The model is found to properly represent the large sale spatial variability of the SST, with colder temperatures around the Gulf of Lions, warmer waters in the Algerian basin south of the Balearic Islands and some relatively colder inflow of Atlantic Water from the Strait of Gibraltar. Finer details associated with mesoscale eddies and filaments do not generally coincide between the free run model and observations. In particular local differences are found in the REP14-MED area, with the model slightly overestimating surface temperatures due to an apparent more pronounced advection of relatively warmer waters from the south-west.

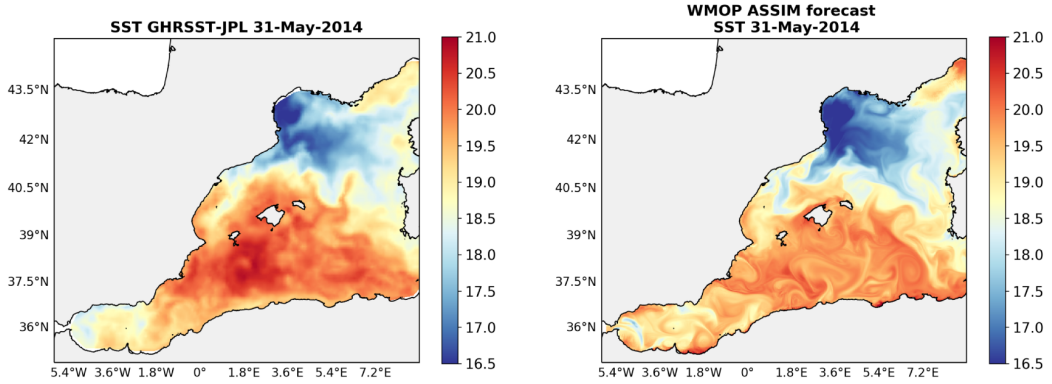

Figure 2.3 *SST for 31 May 2014 from 1) left: GHRSST JPL MUR satellite-derived product, 2) free run WMOP model .*

2.2.3 Data Assimilation System

The WMOP data assimilation system is based on a Local Multimodel Ensemble Optimal Interpolation (EnOI) scheme. It consists in a sequence of analysis (model updates given a set of observations) and model forward simulations. During the analysis step, the state vector $\mathbf{x}^a$ is updated according to equation (2.1), where $\mathbf{x}^f$ is the background model state vector, $\mathbf{H}$ is the linear observation operator projecting the model state onto the observation space and $\tilde{\mathbf{K}}$ is the Kalman gain estimated from the sample covariances (eq. 2.2). $\mathbf{y}$ is the vector of observations. Matrices $\tilde{\mathbf{P}}^f$ and $\mathbf{R}$ are the error covariance matrices of the model and the observations, respectively.

$$\mathbf{x}^a = \mathbf{x}^f + \tilde{\mathrm{K}}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \tag{2.1}$$

$$\tilde{\mathbf{K}} = \tilde{\mathbf{P}}^f \mathbf{H}^T (\mathbf{H}\tilde{\mathbf{P}}^f \mathbf{H}^T + \mathbf{R})^{-1} \tag{2.2}$$

$\tilde{\mathbf{P}}^f$ contains the background error covariances (BECs) estimated by sampling three long-run simulations of the WMOP system with different initial/boundary forcing (coming from CMEMS MED and GLOBAL models) and momentum diffusion parameters. More concretely, for each analysis, a 80-realization ensemble is generated to calculate the BECs. Ensemble realizations are multivariate model fields sampled from the three simulations during the same season, with a time window of 90 days centered on the analysis date. The seasonal cycle is removed from the ensemble anomalies to discard the corresponding large scale correlations mainly affecting temperature. Following this procedure, the computed BECs reflect the spatial variability and anisotropy of the ocean mesoscale circulation. They also represent dynamically consistent covariances between different model variables and depths. Moreover, a domain localization strategy (Ott et al., 2004) is used to dampen the impact of remote observations. A 200km localization radius is used, determined by both the size of mesoscale structures and the approximate distance between two Argo platforms in the Western Mediterranean basin. Here, the domain localization consists in computing independent analysis for each water column of the WMOP domain, considering only the observations located within a 200km radius. It allows to locally dampen the impact of remote observations in the presence of spurious long-range correlations. The code used in this study is written in C and is an adaptation of the EnKF version used in Mourre et al. (2006b), Mourre et al. (2006a) and Mourre and Chiggiato (2014). It was also previously used during the Alborex experiment carried out in the Alboran Sea (Pascual et al., 2017).

In the EnOI, as in any other sequential data assimilation scheme, a special care needs to be brought to the model initialization after analysis updates (Oke et al., 2008). When restarting the simulation from an analysis field, the multivariate initial fields may be violating some physical constrains, such as mass conservation. The model response to balance this state may generate some spurious waves and introduce noise into the system. To minimize such effects, a nudging strategy has been implemented to restart the model after the analysis. In concrete terms, after an analysis is computed, the model is restarted 24 hours before the analysis date applying a strong nudging term in the temperature, salinity and sea level equations towards analyzed values. The time scale associated with this nudging term is 1 day. The nudging procedure reduces the model correction, but guarantees updated multivariate fields closer to the model equation balances, which limits instabilities.

The assimilation system implemented here uses a 3-day cycle (Figure 2.4), which was determined according to the time spent by the gliders to complete one zonal transect. All measurements collected during these 3 days are considered synoptic in the data assimilation process. Altimeter sea level anomalies (SLA), satellite sea surface temperature (SST) and Argo temperature and salinity profiles are assimilated over the whole WMOP domain. For each analysis, a 5-day time window in the past is defined to select Argo observations. This window corresponds to the interval between two profiles provided by a single platform. This ensures that every model point is bounded by at least one Argo profile within the 200km localization radius during each analysis. Concerning altimetry, the last 72-hour CMEMS along-track reprocessed fil-

tered sea level observations are considered for the analysis. The SST field is given by the daily L4 near real time GHRSST JPL-MUR satellite-derived interpolated product (https://podaac.jpl.nasa.gov/dataset/JPL-L4UHfnd-GLOB-MUR). The last available field before analysis is considered. The original 1-km resolution data is smoothed and interpolated onto a 10km-resolution grid to limit the number of observations considered for each analysis. The selected resolution is considered to be sufficient to represent the main circulation features and mesoscale structures present in this SST product, permitting at the same time an affordable computational cost.

The glider profiles are considered as vertical. The corresponding observations are binned vertically and a single value is given for each model grid cell. The representation error is the addition of a vertical and a horizontal components. For each vertical level, the observed variance in the vertical grid cell is used as an approximation of the vertical representation error. In addition, the horizontal representation error variance is assumed to be $(0.25°C)^2$ and 0.05^2 for temperature and salinity (in practical salinity scale) measurements, respectively. The choice of the variance in the vertical grid cell as a representation error intends to represent the possible variability due to the consideration of only one observation for the whole grid cell depth, due to the binning process. The horizontal representation error accounts for possible errors due to unresolved processes or scales not represented in the model. We then add these two variances, as we consider the horizontal and vertical components as independent. CTD observations are binned vertically in a similar way before assimilation, considering the representation error in an analogous way.

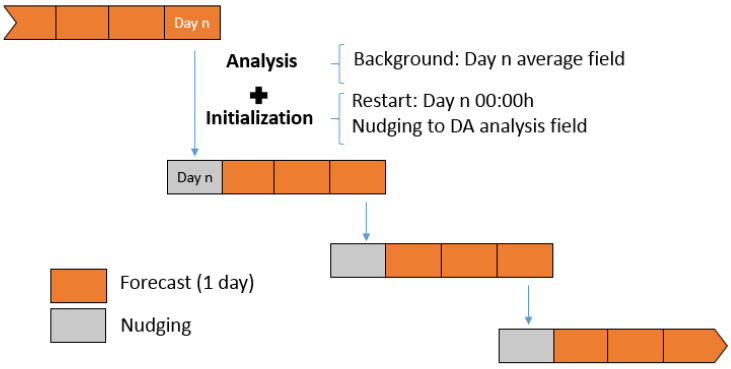

Figure 2.4 *Scheme of the 3-day data assimilation cycles.*

Figure 2.5 illustrates the innovations (differences between the observations and the background model) computed for the first analysis carried out on 31 May. It shows the absence of significant biases in the model prior to data assimilation, which is a prerequisite for an effective assimilation of the observations. Moreover, the magnitude of the standard deviation of innovations of the surface variables (0.42°C for SST and 0.042m for SLA) properly matches that of the observation error (0.56°C for SST and 0.036m for SLA) and the ensemble spread (1.10°C for SST and 0.056m for SLA for the analysis on 31 May), which guarantees the necessary overlap between the probability density functions of model and data.

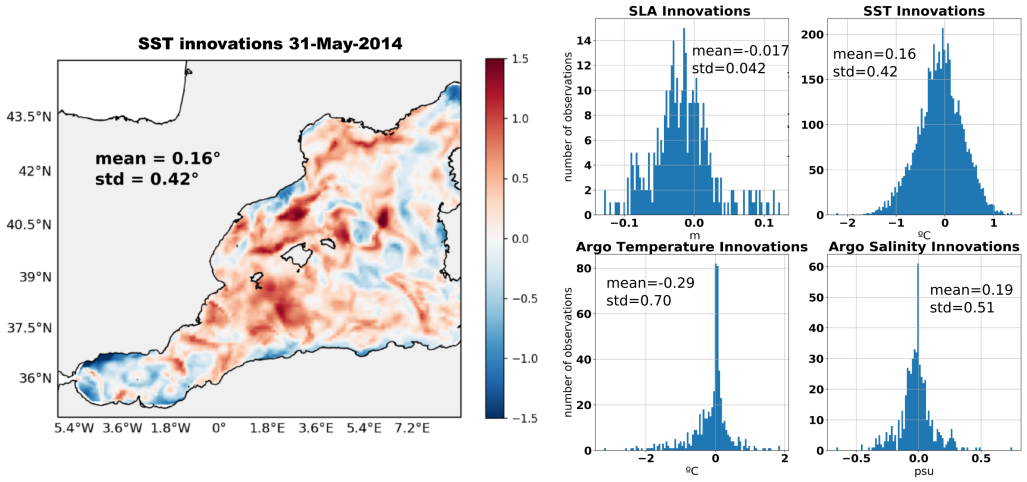

Figure 2.5 *Left: SST misfits between the observations and the free run model on 31 May 2014. Right: histograms of the innovations for the different sources of observation ingested by the assimilation system. The corresponding mean and standard deviation are provided in each panel.*

2.2.4 Experiments

In addition to the background simulation without any data assimilation (hereafter NO_ASSIM), seven simulations were produced spanning the period 1-24 June 2014, assimilating different sets of observations. The first simulation (GNR) assimilated generic observations from satellite along-track SLA, satellite SST and Argo temperature and salinity profiles. The second simulation (GNR_CTD) assimilated these generic observations plus all CTD temperature and salinity profiles collected during Leg 1. The five remaining simulations assimilated the generic observations plus glider temperature and salinity data from one to eight vehicles (GNR_1G, GNR_2G, GNR_3G, GNR_4G, GNR_8G), selected among the available platforms to optimally cover the area of interest. For example, GNR_1G considers the glider which travels in the center part of the domain, GNR_2G selects the two gliders which divide the study region in 3 areas of similar dimensions, etc... The different sets of vehicles selected for these simulations are illustrated in Figure 2.6.

The whole timeline of the numerical experiments is described in Figure 2.7. A spinup period of 9 days was imposed for all these data-assimilative simulations, during which only the generic observations were assimilated. As explained previously, the 3-day assimilation cycle implemented in this study corresponds to the time spent by a glider to complete a zonal transect. In the case of the CTDs, as the duration of the data collection in Leg 1 was six days, the data was assimilated during two cycles.

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

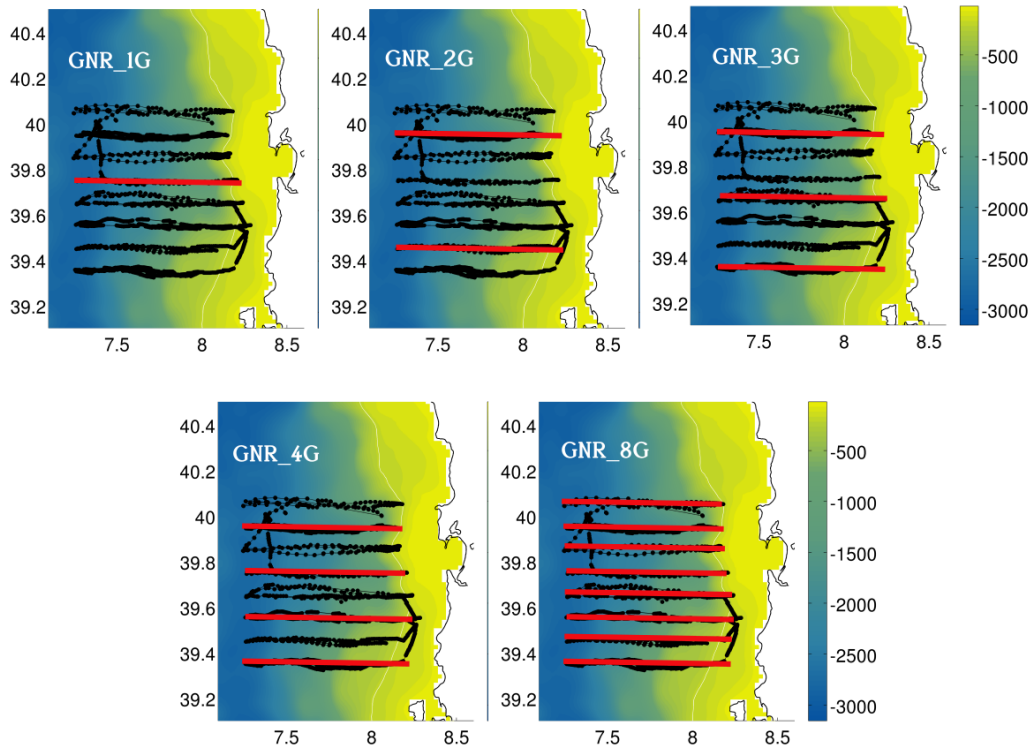

Figure 2.6 *Illustration of the different sets of gliders selected in the different data-assimilative experiments. The position of all glider measurements are shown in black dots. The red zonal lines indicate the selected glider tracks in each of the experiments. The name of the corresponding simulation is specified in each panel.*

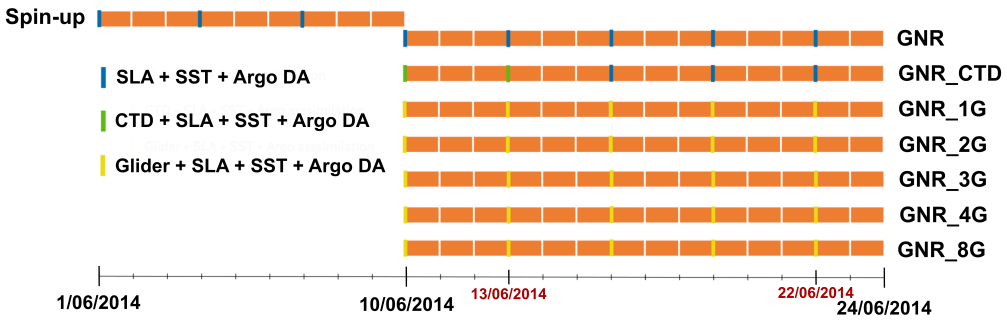

Figure 2.7 *Timeline of the seven data-assimilative simulations. The analysis dates are highlighted in color, indicating the assimilated datasets.*

2.3 Results

We evaluate in this Section the performance of the data assimilation following three successive steps. We first verify that the data from the different sources are properly ingested in the system over the whole modelling area both during the spinup period and subsequent assimilation phase. Then, we examine the impact of the assimilation of the local and dense observations datasets onto the temperature, salinity and density fields in the REP14-MED area. Finally, we assess the performance of the simulations against

independent data from CTDs and Scanfish observations collected during Leg 3.

2.3.1 Data ingestion and performance over the whole modelling area

We first assess here the performance of the assimilation during the spinup period by analyzing the evolution of the Root-Mean-Square-Difference (RMSD) between the model simulations and satellite SLA, SST observations and Argo profiles. For each source and variable, the RMSD is calculated as expressed in Equation 2.3 below, where o_i and m_i take the values of the observations and their model equivalents, respectively. n is the number of observations. To better highlight relative simulation improvements, the RMSD for each specific day is normalized by dividing the RMSD (Eq. 2.4) by that of the simulation without any data assimilation for that specific day. A reduction of the normalized RMSD indicates that the analyzed field is closer to the observations than the background field without assimilation.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n}(o_i - m_i)^2}{n}} \tag{2.3}$$

$$RMSD_{normalized} = \frac{RMSD_{assim\_simulation}}{RMSD_{no\_assim}} \tag{2.4}$$

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

The results are presented in Figure 2.8. They show a satisfactory and continuous reduction of the RMSD for all the sources of data and variables, indicating a good system performance. The normalized RMSD is significantly reduced during the first analysis (between 20 and 60% depending on the analyzed variable), it then tends to slightly increase until the next analysis 3 days later, which reduces it again in most of the cases. In some occurrences, the RMSD continues decreasing during two days after the analysis. The overall persistence of the correction between two assimilation dates is especially remarkable. It reveals the general proper performance of the assimilation system, which is able to recursively correct the multivariate fields without introducing spurious structures and instabilities which would significantly alter the system.

After this verification of the overall satisfactory performance of the data assimilation during the spinup period in terms of RMSD, the same kind of assessment was performed for the seven subsequent simulations, the GNR control simualtion and the ones assimilating either CTDs or glider observations during the field experiment besides the generic observations. We only show here results from the GNR_CTD simulation (Figure 2.9), since the behaviour is very similar for the rest of the simulations. The system still properly reduces the normalized RMSD in terms of SST, SLA and T-S profiles at Argo locations, with similar reductions as that observed during the spinup period (from 20 to 60% of error reduction).

An important aspect in this comparison is that the assimilation of high resolution T-S

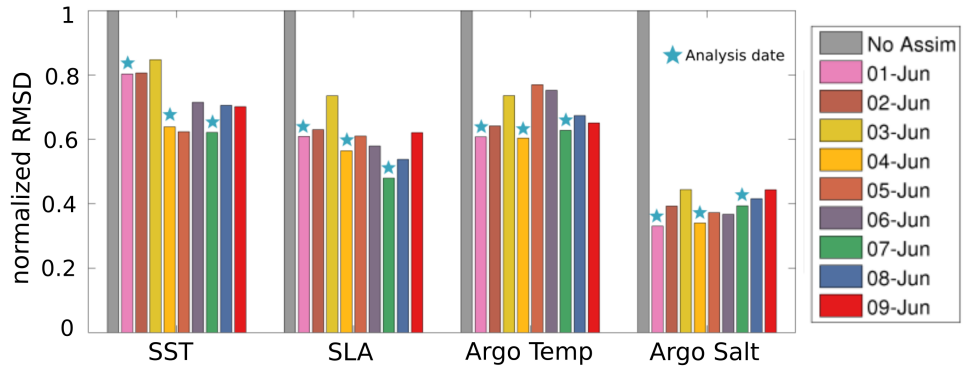

Figure 2.8 *Evolution of the normalized RMSD against observations for the spinup simulation.*

profiles data in the REP14-MED area does not negatively affect the overall performance of the system over the whole modelling area. This could happen through the generation of spurious structures in the densely observed area which could then propagate over the domain. Moreover, the reduction of the normalized RMSD with respect to CTD observations shows that the local observations have also been properly ingested in the system. Notice that the relatively larger SLA RMSD found during the period 10-23 June compared to the spinup period also affects the GNR simulation. Therefore, it is not due to the incorporation of CTD observations, but rather related to the natural evolution of SLA errors.

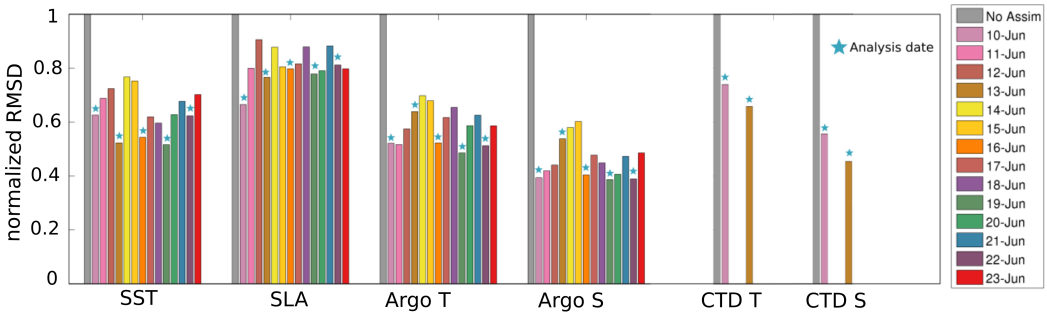

Figure 2.9 *Evolution of the normalized RMSD against observations for the GNR_CTD simulation.*

2.3.2 Temperature, salinity and density fields in the REP14-MED area

To complement these statistical diagnostics based on the normalized RMSD, we analyze here the temperature and salinity fields in the REP14-MED trial area on 13 June. This corresponds to the first day after the second analysis of CTD and glider data assimilation cycles. At that time, either all CTD data from Leg 1, or one back-and-forth transect from the gliders have been introduced into the system. Figure 2.10 shows temperature, salinity and potential density daily average fields at 50 m depth for four of the simulations (NO_ASSIM, GNR, GNR_CTD and GNR_8G) on 13 June. The temperature and salinity data assimilated until that date are also represented as colored dots on the panels corresponding to GNR_CTD and GNR_8G.

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

The additional assimilation of dense local data from CTDs and gliders further modulates these patterns, producing smaller scale patches and filaments. Two main warm temperature anomalies are detected in both CTD and glider observations at 50m depth, associated with relatively slightly salty anomalies. The strongest one, located around 7.7E-39.3N, is somehow represented in both simulations GNR_CTD and GNR_8G, with a more pronounced signature in GNR_8G. Notice that this signature is not fully coincident with the observed location displayed in Figure 2.10 due to the evolution of the model from the first analysis on 10 June to the time of the plot 3 days later. The second relatively warm temperature patch found around 7.5E-40N is less marked that the first one. Again, it is somehow better reproduced in GNR_8G than GNR_CTD. The observations, from both CTDs and gliders, are characterized by an energetic small scale variability, which translates into small scales and filamental structures in the model after data assimilation. Notice also the improvements in the relatively higher salinity along the coast after assimilation of the measurements from the CTDs and the gliders. The relatively high salinity patch around 39.5N seen in the GNR simualtion is strongly attenuated in both GNR_CTD and GNR_8G. The density fields of GNR_8G exhibit two areas of lower potential density associated with these two anomalies, in good qualitative agrement with the observations even if the magnitude of the gradients is reduced compared to the measurements. These density anomalies are less clear in GNR_CTD. Both simulations show denser water on the north-eastern part of the domain and similar overall circulation patterns which significantly differ from NO_ASSIM and have also marked differences with GNR. A common property observed in both simulations is the current flowing northeastwards in the central part of the sampled area and bifurcating near the coast, with one branch directed southwards and other northwards, giving also rise to cyclonic and anticyclonic eddies with dimensions around 30-40km.

2.3.3 Performance assessment using independent data during Leg 3

As a third step we analyze here the realism of the simulations during Leg 3 using independent observations which have not been assimilated in the experiments. More specifically, we compared the model outputs on 22 June (after all assimilation cycles have been completed) with CTD and Scanfish temperature and salinity observations collected between 20 and 23 June. A qualitative analysis is first performed, based on the potential density fields reconstructed from both CTD and Scanfish observations at 50m depth. The DIVA software (Data Interpolating Variational analysis, (Barth et al., 2010)) and its web interface (http://ec.oceanbrowser.net/emodnet/diva.html) have been used to generate the interpolated density field. Figure 2.11 compares the density fields at 50 meter depth from the different simulations with this density field derived from the observations. Model currents at that depth are shown as well as the CTD and Scanfish observations.

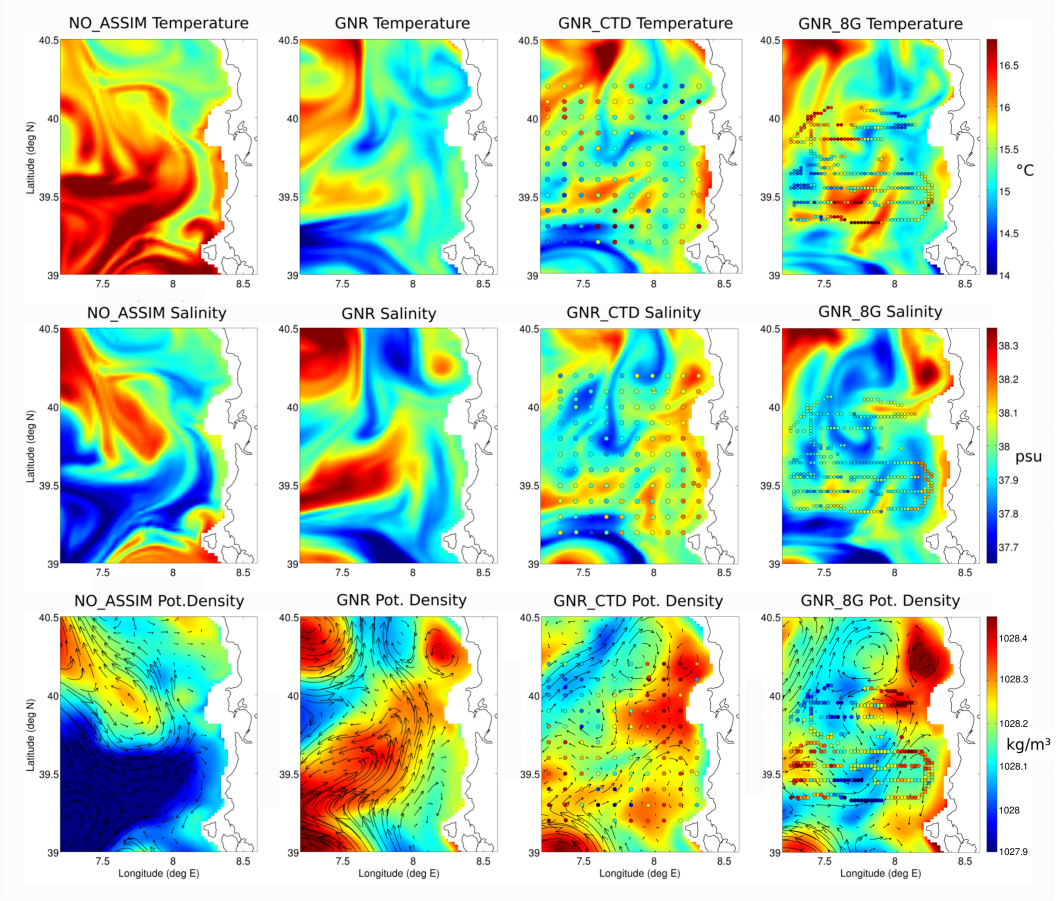

Figure 2.10 *Model temperature (upper panels), salinity (middle panels) and potential density and currents (lower panels) at 50m depth on 13 June. From left to right: simulations NO_ASSIM, GNR, GNR_CTD and GNR_8G. The assimilated data are superimposed as colored dots in the temperature and salinity panels for the two simulations GNR_CTD and GNR_8G.*

The main features represented in the reconstructed density field derived from the observations include a marked negative density anomaly centered around 7.8E-39.4N with a spatial extension around 40km, a coastal fringe with relatively denser waters, and a second patch of denser water between 39.5N and 40N on the Western side of the domain. These features were somehow already present during Leg 1 (see Figure 2.10 and Section 3.2).

All the data-assimilative simulations represent the denser coastal fringe and the associated southward flow, yet with different characteristics. It extends offshore, associated with a cyclonic eddy, in GNR, GNR_CTD, GNR_1G and GNR_2G. GNR_4G and GNR_8G qualitatively provide a more accurate shape of this coastal feature. In addition, these two simulations better represent the secondary relatively denser patch on the western side. Lower density anomalies south of 39.5N are also present in all the simulations. GNR, GNR_CTD and GNR_1G seem to qualitatively better match the reconstructed field by representing an anticyclonic eddy around a local density minimum, with an approximate 40km diameter. However, the exact shape of this anomaly and in particular its meridional extension, was not properly observed during Leg 3, which only provided a single Scanfish zonal section at 39.4N across this anomaly. While it is rep-

resented as a close eddy in the reconstructed field due to the interpolation method, the more elongated shape in the meridional direction provided by GNR_4G and GNR_8G is also consistent with the Scanfish observations. Notice that the data-assimilative simulations all qualitatively improve the solution without data assimilation. Among them, GNR_4G and GNR_8G provide a particularly remarkable pattern agreement with the Scanfish and CTD observations.

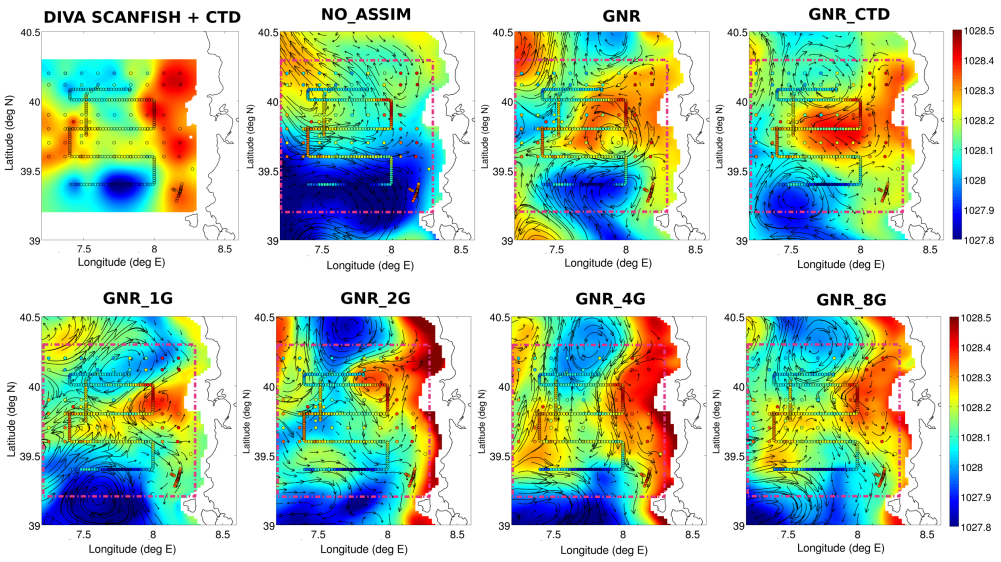

Figure 2.11 *Top left panel: potential density field reconstruction from Scanfish and CTD data collected between 20 and 23 June ($kg \cdot m^{-3}$). Remaining panels: potential density ($kg \cdot m^{-3}$) and model currents at 50m depth on 22 June for the seven simulations NO_ASSIM, GNR, GNR_CTD, GNR_1G, GNR_2G, GNR_4G, GNR_8G and GNR_CTD.*

To quantify the improvement, we now present the normalized RMSD, both considering the type of large scale observations which were assimilated over the whole domain, and the independent sea trial observations used in this Section. We computed the normalized RMSD for each of the seven data-assimilative simulations on 22 June and for the different sources of observation (Figure 2.12). CTD and Scanfish observations between 20 and 23 June were considered synoptic for this purpose.

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

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

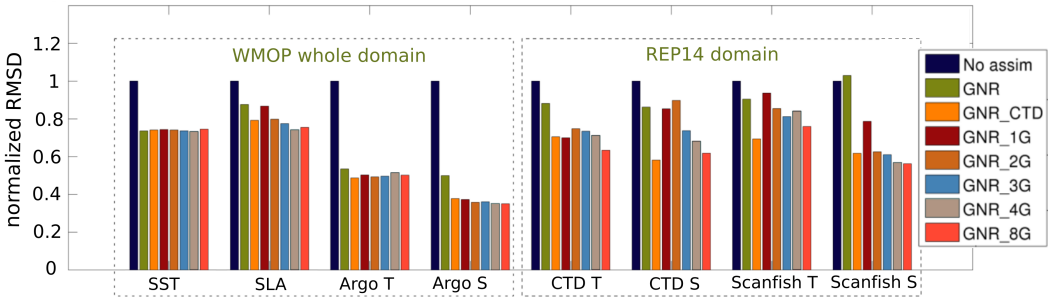

Figure 2.12 *Normalized RMSD against observations on 22 June for the 7 numerical simulations. Dashed bounding boxes delimitate on the one side the observations assimilated over the whole domain, and on the other side the independent campaign observations within the REP14-MED domain.*

observations also reduces the RMSD, with an overall enhanced performance as the number of platforms increases. The comparison with different platforms and variables provides slightly different rankings of the simulations. For instance, in this comparison, GNR_1G provides a similar performance as GNR_CTD against independent CTD temperature data, but a lower performance against CTD salinity and Scanfish measurements. The assimilation of data from 4 gliders improves the performance with respect to the assimilation of CTDs when comparing to Scanfish salinity data, but the performance is lower when comparing to the other sources. GNR_CTD provides the best RMSD reduction when comparing to CTD salinity and Scanfish temperature, but it is GNR_8G which shows the best performance when considering CTD temperature and Scanfish salinity data. These variations are probably due to the specific spatial sampling of the CTDs and Scanfish (see Figure 2.2) combined to the high spatial oceanic variability in the area.

An average RMSD reduction number is obtained here by computing the square root of the average normalized mean square difference over the four comparisons (CTD and Scanfish temperature and salinity) in the REP14-MED domain. These synthetic average normalized RMSD scores are presented in Figure 2.13. This average RMSD gives scores of 39% and 40% of error reduction for the simulations GNR_CTD and GNR_8G, respectively. According to this overall metric, the CTD survey is more performant than a sampling using 4 gliders, and shows very close performance as that obtained with 8 gliders. Notice that the average normalized RMSD illustrates the progressive gain obtained when increasing the number of gliders considered in these experiments.

2.4 Discussion

The assimilation of observations is crucial to improve forecasts. Regional forecasting systems should be able to efficiently combine high resolution local profile data and larger scale satellite observations over an extended modelling domain. The recursive Ensemble Optimal Interpolation scheme employed in this study is shown to be able to ingest both types of data and to systematically reduce the errors when compared to a control free-run simulation. Even if the EnOI is theoretically inferior to more advanced data assimilation schemes such as the ensemble Kalman filter or the 4DVar (see for instance Sakov and Sandery (2015) and Oke et al. (2007) for comparisons of DA

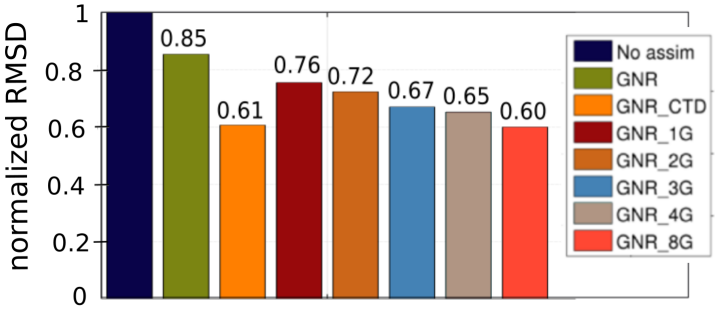

Figure 2.13 *Average normalized RMSD against independent observations in the REP14-MED areas on 22 June for the 7 numerical simulations.*

schemes), its numerical efficiency makes it a good compromise for operational and practical implementations with high resolution models.

The domain localization approach, which does not take into account the observations further than a given radius to correct the field at a given location, guarantees that the assimilation of dense profile observations from gliders and CTDs over a reduced area does not degrade the results over the whole modelling domain. Moreover, it allows to reduce the RMSD and to improve the representation of local water masses and the associated circulation in the reduced REP14-MED area which has dimensions around 100km. The corrections introduced by the assimilation of CTD data during Leg 1 are found to remain in time, providing a very positive and significant error reduction when comparing to independent measurements 10 days after the initial CTD data collection.

The error reduction in the salinity when compared with Argo profiles is relatively higher than for the temperature. We believe that this is the result of two different effects. First, the influence of assimilating SST satellite observations has a high impact on the top layers of the model. As a consequence, the model does not fit the Argo temperature profile observation so closely, which makes discrepancies higher. Although the covariances are multivariable in our system, for a temperature observation the correlations are higher with the temperature field than with the salinity field in the sorrounding area and in the vertical. Consequently, the impact of SST observations on the salinity fields is lower than in temperature ones. Secondly, the ratio between the observation error and the ensemble spread is significantly lower for the salinity, at specific depths where we found high values of innovations. This higher spread is due to the multimodel strategy we employ for the ensemble generation. The mean salinity profile, computed for different regions of the domain, has some differences in depth for the different hindcast simulations, making the ensemble have a relatively high spread at those levels. This implies that the analysis is closer to the observations, leading to a better correction of the misfit between the Argo floats and the model profiles. In the figure 2.14 below we can observe the ensemble standard deviation for temperature and salinity fields at 300 m depth for a specific analysis date, together with the position of the floats assimilated during the whole simulation. In many of the points where Argo floats have been assimilated, the standard deviation for temperature and salinity are of the same order, while the errors considered for the observations are an order of magnitude lower for salinity (0.05^2) than for temperature observations (0.25^2).

As a limitation, we notice that the oceanographic structures of small horizontal and

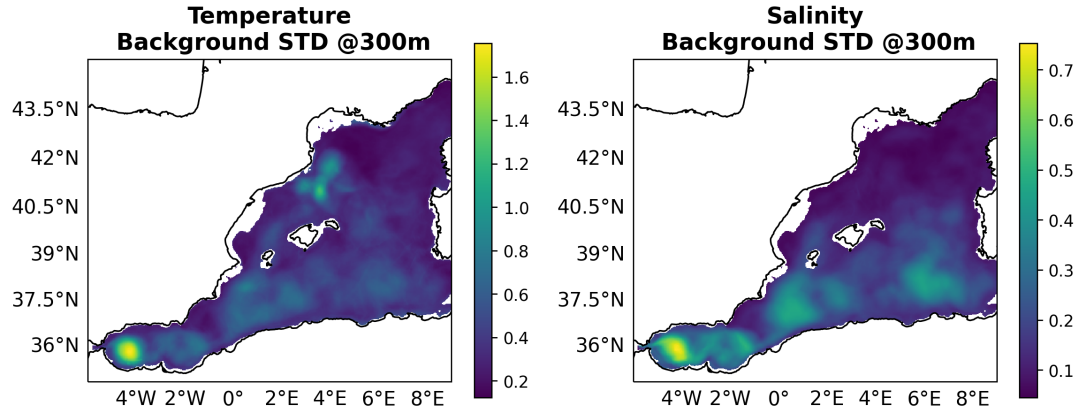

Figure 2.14 *Ensemble forecast standard deviation for temperature (left) and salinity (right) fields at 300m. Red dots represent the position of the Argo floats assimilated along the simulation.*

vertical dimensions, which have a strong signature in the dense observation datasets, are only approximately represented in the temperature and salinity fields just after assimilation, as shown in Figure 2.10). This is the case for instance for the strong positive temperature anomaly around 7.7E-39.3N, which is the signature of an eddy with an horizontal diameter around 40 km and a vertical dimension around 50m. We attribute this limitation on the one hand to the smoothing effect of the background error covariances, which impacts both along the horizontal and vertical directions, and on the other hand to the nudging initialization procedure, which attenuates the model correction with the aim to provide more dynamically consistent fields. To illustrate the error covariances of our EnOI implementation, Figure 2.15 shows both the horizontal and vertical model ensemble correlations generated from the ensemble for the analysis on 22 June.

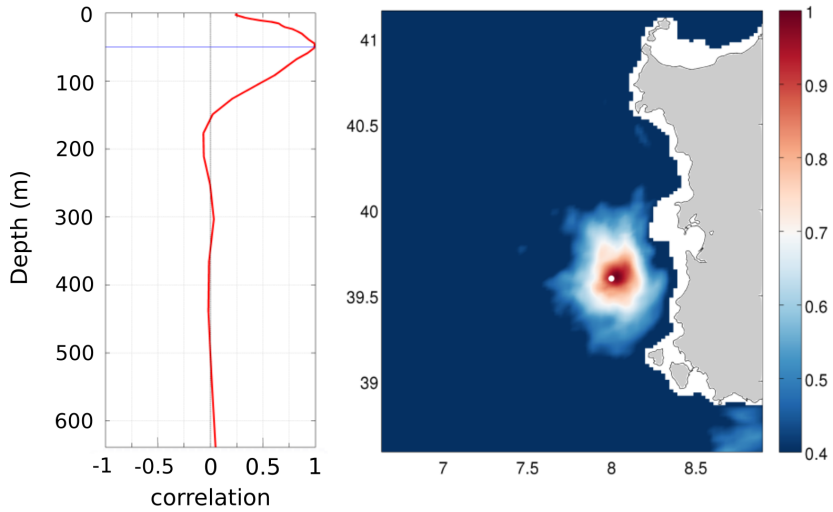

Figure 2.15 *EnOI temperature ensemble correlations for a temperature observation within the REP14-MED domain at 50m depth (position indicated by the white dot on the right panel). Left: correlations along the vertical. Right: horizontal correlations at 50m depth.*

A spatial smoothing takes place during the assimilation, affecting the area with significant correlations with the observed locations. The ensemble correlation distances are found to be around 100m in the vertical and 75km in the horizontal, being then larger than the smaller scale structures observed in the CTD and glider surveys. Two-steps assimilation strategies separating long- and short-distance correlation scales might allow to improve the representation of these finer patterns in more sophisticated data assimilation systems (e.g. Li et al. (2015)).

The second factor explaining this limitation is related to the nudging initialization strategy, which has the advantadge of limiting undesired model shocks after the analysis, but also attenuates the corrections and therefore the agreement with observations. The simple nudging technique used in this work is easy to implement and cost-effective. It could be improved in the future by considering more advanced approaches (e.g. Sandery et al. (2011)).

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

While CTDs allow a relatively fast comprehensive description of a specific study area, gliders provide a slower sampling but also allow a repetition of specific monitoring tracks over a longer period. In this study, the CTD initialization survey results in a similar forecast performance after data assimilation in terms of RMSD reduction as an 8-glider continuous monitoring of the area flying along predefined paths with regular spacing. It should be highlighted that the meridional spacing in the case of the 8 gliders fleet is the same as for the CTD casts (10km). The improvement provided by the higher spatial resolution offered by gliders in the zonal direction might be limited by the spatial resolution of the model, which do not allow to ingest the very fine-scale features observed by the gliders. In that sense, it is likely that glider data assimilation would further benefit from an increase of the model resolution. Let's mention that while glider platforms are considered autonomous, their operation still implies a very significant effort in terms of platform deployment, recovery, piloting and maintenance. The models could also highly benefit from the near real-time controllability of gliders, allowing to continuously adjust their path along optimal routes in the study area. In this framework, efficient adaptive sampling procedures should theoretically allow to use a reduced number of gliders to reach the same level of performance (Lermusiaux, 2007) and lead to a better description of specific targeted features, as long as their representation is permitted by the model resolution. The definition of optimal collective behaviours based for instance on glider fleet coordination or cooperation (e.g. Alvarez and Mourre (2014)) also constitutes an interesting field of research in that direction.

2.5 Conclusions

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

tion survey or a glider fleet sampling, in improving model forecasting capabilities in a specific area.

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

While the assimilation of generic observations from SST, SLA and Argo leads to an average error reduction of 15% when comparing to independent measurements collected during Leg 3 of the sea trial in the REP14-MED area, the assimilation of glider and CTD data allows an additional significant improvement. Gliders, which provide a continuous sampling of the area along regularly spaced zonal tracks, allow to reduce the forecast error as the number of platforms increases. The consideration of one glider leads to a 24% average error reduction with respect to the simulation without assimilation. This percentage increases to 28%, 33%, 35% and finally 40% for the 2-glider, 3-glider, 4-glider and 8-glider fleet configurations, respectively. Incrementing the number of gliders results in a better representation of the ocean state captured by observations, with a most accurate representation of the mesoscale structures and associated circulation.

The assimilation of the observations from the dense initialization survey based on 10-km spaced CTD stations leads to an average error reduction of 39%: it outperforms the 4-glider configuration and provides very similar results in terms of RMSD as the 8-glider fleet configuration. The 10km spacing offered by both sampling strategies is essential here to improve the representation of the mesoscale variability in the study area. In view of these results, gliders certainly provide a very interesting alternative to traditional CTD surveys used to initialize high-resolution regional ocean models, provided that a fleet of vehicles can be deployed at sea. Moreover, an increased performance can certainly still be expected by optimizing the regular track sampling carried out in this experiment through adaptive sampling procedures.

Chapter 3

Evaluating High-Frequency radar data assimilation impact in coastal ocean operational modelling

Abstract

The impact of the assimilation of HFR (High-Frequency Radar) observations in a high-resolution regional model is evaluated, focusing on the improvement of the mesoscale dynamics. The study area is the Ibiza Channel, located in the Western Mediterranean Sea. The resulting fields are tested against trajectories from 13 drifters. Six different assimilation experiments are compared to a control run (no assimilation). The experiments consists in assimilating (i) Sea surface temperature, sea level anomaly and Argo profiles (generic observation dataset); the generic observation dataset plus (ii) HFR total velocities and (iii) HFR radial velocities. Moreover, for each dataset two different initialization methods are assessed: a) restarting directly from the analysis after the assimilation or b) using an intermediate initialization step applying a strong nudging towards the analysis fields. The experiments assimilating generic observations plus HFR total velocities with the direct restart provides the best results, reducing by 53% the average separation distance between drifters and virtual particles after the first 48 hours of simulation in comparison to the control run. When using the nudging initialization step, the best results are found when assimilating HFR radial velocities, with a reduction of the mean separation distance by around 48%. Results show that the integration of HFR observations in the the data assimilation system enhances the prediction of surface currents inside the area covered by both antennas, while not degrading the correction achieved thanks to the assimilation of generic data sources beyond it. The assimilation of radial observations benefits from the smoothing effect associated with the application of the intermediate nudging step.

3.1 Introduction

High-frequency radars (HFR) are a fast-growing technology, playing an important role in coastal observing systems around the world (Roarty et al., 2019; Rubio et al., 2017b). They allow real-time measurements providing a new, detailed, and quantitative description of physical processes at the marine surface (Paduan and Washburn, 2013). Their capacity to measure currents at high spatial and temporal resolution over relatively

large coastal areas make them a convenient system for operational purposes. They can be used to validate numerical models (Aguiar et al., 2020; Mourre et al., 2018; Lorente et al., 2021a), analyze Lagrangian dynamics (Hernández-Carrasco et al., 2018) or constrain numerical models via data assimilation (DA) (Vandenbulcke et al., 2017; Iermano et al., 2016b; Janeković et al., 2020b).

HFR is a cost-effective shore-based remote-sensed technology exploiting the Bragg resonance phenomenon (Crombie, 1955) to map ocean surface currents, wave fields, and increasingly winds, in coastal areas. They complement satellite altimeter observations which are limited to larger scales and suffer limitations when approaching the coast (Vignudelli et al., 2019; Pascual et al., 2013). The capability of HFR to give realistic observations of surface currents has been widely validated (Chapman et al., 1997; Emery et al., 2004; Paduan et al., 2006). Furthermore it has been used to validate geostrophic currents computed from along-track altimetry (Pascual et al., 2015) and to correct sea surface height (SSH) altimeter fields (Roesler et al., 2013).

Regional ocean models are invaluable tools for operational oceanography (Wilkin and Hunter, 2013; Onken et al., 2008; De Mey-Frémaux et al., 2019). However, they are inevitably affected by errors from multiple sources, especially in highly dynamic coastal areas where conditions tend to change rapidly. In such places, assimilation of observations such as the ones provided by HFR can help to constrain the model solution and improve the forecast. Assimilation of HFR data has been successfully applied in different regions around the globe, starting from the pioneer study conducted by Breivik (2001) using an optimal interpolation (OI) scheme. Since then, many different studies have explored the performance of HFR for data assimilation in ocean circulation models, using different assimilation schemes, data types and techniques. Authors have employed high-frequency (hourly) or filtered data depending on the focus of the study and the dynamical processes of interest. The use of whether radial or total observations has also been object of study and debate. While radial velocities theoretically provide a larger amount of information without any data processing, they are prone to present higher spatial gradients than reconstructed total observations, which incorporate some spatial smoothing. Besides, radials have a wider coverage, providing data in areas only covered by one antenna and can be used even in case of failure of the other antennas. Theoretically and under the assumptions of linearity and normal distribution of errors in the state dynamics and measurements, as well as in the transformation from radials to totals, the assimilation of radial currents should outperform the assimilation of total currents, since all the information of the totals is included in the radials and the later contain additional information which is not included in the totals. However, in real-world experiments, these major assumptions are not verified. In particular, the model is non-linear, observation error covariances are not Gaussian and certainly not perfectly known and the transformation from radials to totals also involves nonlinearities. In the literature, both kinds of observations have been assimilated with satisfactory results. To our knowledge, Shulman and Paduan (2009) is the only work evaluating the contribution of both radial and total velocities in the same experiment. Their results show the capacity of the system to improve surface currents and circulation down to 120m depth in areas covered by two or more antennas, for both kinds of data. Depending on the position of the mooring with respect to the coverage of the antennas, their validation showed a varying complex correlation against mooring observations when using radial or total observations. Using observations from only one antenna, Shulman and Paduan (2009) found that results were extremely variable and highly dependent on the direction

of the bearing with respect to the dominant flow.

Oke et al. (2002) used an OI scheme to assimilate low-pass filtered surface total velocity measurements from an HFR array to correct model circulation. They used a so-called TDAP (Time-Distributed Averaging Procedure) initialization method after analysis, which progressively applies data assimilation increments, to preserve appropriate dynamical balances. This data assimilation approach resulted in an increase of the correlation between model and observations from 0.42 to 0.78.

More recently, hourly reconstructed total currents have also been employed using both sequential (Ren et al., 2016; Paduan and Shulman, 2004) and variational data-assimilation schemes such as 4D-Var (Zhang et al., 2010; Wilkin and Hunter, 2013; Yu et al., 2016). However, depending on the model set-up and the oceanic processes of interest, the use of hourly data may not be the most appropriate, as for instance in Kerry et al. (2016), where radial speeds and angles are spatially averaged onto the model grid and a 24 h boxcar-averaging filter is used to remove tides and inertial oscillations that are not resolved by the model. Kerry et al. (2018) show that among all the assimilated observations, HFR were the ones which had the larger impact on the currents and the transport in the Eastern Australian Current.

The use of hourly data in sequential data assimilation schemes is not straightforward, due to the analysis frequency which is generally larger than one hour. An option is to use an extended state vector as in Barth et al. (2008), who employed an ensemble based Kalman Filter (KF) method using hourly radial observations in the West Florida Shelf. For the initialization Barth et al. (2008) implemented a spatial filter and averaged the ensemble fields in an attempt to remove spurious variability before it is introduced into the model. Barth et al. (2011) and Marmain et al. (2014) employed a similar approach, using all radial hourly observations available during the assimilation window and an extended state vector to correct the wind forcing fields and boundary conditions respectively in a similar way to variational methods. While Barth et al. (2011) showed that the correction had a positive impact on the reconstructed winds and the SST in the German Bight, Marmain et al. (2014) found an improvement in surface currents in the North-Western Mediterranean Sea, although with some degradation on the density fields and under surface currents. Stanev et al. (2015) also used hourly radial observations to correct tidal currents in the German Bight. In an operational context and based on a spatio-temporal optimal interpolation (STIO), Stanev et al. (2016) demonstrated that their system had a good skill to correct currents even beyond the HFR covered area.

A comparison of the impact of both time-filtered and unfiltered HFR currents (with respect to a model with and without tides) was done in Shulman and Paduan (2009), showing that the sub-tidal period velocity simulations were similarly improved through the assimilation of either low-pass-filtered surface currents or instantaneous (hourly) surface currents. More recently, Vandenbulcke et al. (2017) using different KF schemes, with an extended state vector, assimilated hourly radial velocities to correct inertial oscillations in a regional model of the Ligurian Sea. They show an important effect on the correction of inertial oscillations during the first 12 hours, when considering all hourly observations in a 48-hour time-window instead of using only the corresponding to one single hour.

In the present study we aim at evaluating the impact in coastal ocean operational modelling of the assimilation of both HFR total and radial velocities, also exploiting different initialization methods after analysis. Our focus is on the correction of mesoscale

structures and larger scale circulation, rather than inertial oscillations or tidal currents.

The study area is the Ibiza Channel (IC, Fig. 3.2), which is the passage between the oriental coast of Spain mainland and the island of Ibiza. It is a crucial area for understanding mixing and transport processes in the Northwestern Mediterranean Sea. Two different water masses interact in the IC: (i) a relatively salty water that has already recirculated in the Western Mediterranean flowing southward along the shelf as the Northern Current, and (ii) a branch of the Modified Atlantic waters transporting fresher waters originally entering through the strait of Gibraltar and flowing northward (Pinot et al., 1994, 1995) on its easternmost part. The dynamics, and the ecological and economical importance of the area have raised a specific interest in understanding the relevant ocean processes (Heslop et al., 2012; Balbín et al., 2014; Pinot et al., 2002; Hernández-Carrasco et al., 2018; Vargas-Yáñez et al., 2021). The analysis of repeated observations along a glider endurance line in the Ibiza Channel has revealed a high variability of meridional transports over time scales of days to weeks (Heslop et al., 2012). This high variability due to the interaction of multiple processes with different water masses over a complex topography make the operational forecasting particularly challenging. The anthropogenic pressure in the region makes it necessary to develop accurate tools for Search and Rescue, oil spill forecasting or larval dispersion to efficiently respond to emergencies and protect ecosystems.

Since 2012, the Balearic Island Coastal Observing and forecasting System (SOCIB, Tintoré et al. (2013)) operates a CODAR HFR system that monitors the IC with two antennas measuring hourly surface currents (Tintoré et al., 2020). This infrastructure is part of the Joint European Research Infrastructure of coastal observatories (JERICO, www.jerico-ri.eu). Lana et al. (2016) validated the IC HFR observations against current-meter, ADCP and surface Lagrangian drifters, showing a good agreement and the absence of significant mean error (hereafter referred as bias). A joint analysis of HFR observations and surface winds in terms of Empirical Orthogonal Functions (EOF) demonstrated that the surface current variability was mainly driven by local winds and mesoscale circulation.

The present research about Ibiza Channel HFR data assimilation was carried out within the Joint Research Action on coastal forecasting of the European Horizon 2020 JERICO-NEXT project. Seven one-month period simulations have been generated to investigate the data assimilation performance of HFR raw radial observations compared to reconstructed totals currents. We have employed three different datasets, and for each of them, two different initialization methods after analysis. Additionally, a free-run simulation without assimilation has been used as control run. An exhaustive assessment has been performed following both Eulerian and Lagrangian approaches, including an independent set of 13 drifters deployed in the area.

The chapter is structured as follows: Section 3.2 describes the data and methods employed, including the DA system and the description of the experiments. Results are presented in Section 3.3. Finally, the discussion of the results and the conclusions are presented in Sections 3.4 and 3.5.

3.2 Data and Methods

3.2.1 High-Frequency Radar

The SOCIB HFR system consists in two CODAR SeaSonde stations of the islands of Ibiza and Formentera (named GALF and FORM, respectively), covering the eastern side of the IC. It has operated since June 2012, providing real-time high-resolution observations of surface currents (Tintoré et al., 2020; Lana et al., 2015, 2016). Each HFR station emits at a central frequency of 13.5 MHz and a bandwidth of 90 kHz, reaching ranges up to 85 km. Emitted electromagnetic waves are back-scattered by surface waves of exactly half the HFR wavelength. Radial velocities (velocities toward or away from the antenna) are derived from the Doppler shift due to the difference between ideal and measured Bragg frequency (Barrick, 2008). At the specified operating frequency, measurement depth is approximately 0.9 m (Stewart and Joy, 1974). Radial observations provide the velocity along a bearing, calculated from radio signals backscattered from the ocean surface. Hourly radial velocity maps from both stations are systematically quality controlled and the total velocity vectors are reconstructed by combining the radial velocities with overlapping coverage, on a regular 3×3 km grid. Each grid point observation is computed using a unweighted least-square fitting (UWLS) (Lipa and Barrick, 1983), considering all radial observations within a 6km radius. Total reconstructed observations have a range up to 65 km off the antenna, compared to the 85 km that radials can reach.

In our experiments, we used daily means HFR observations (Figure 3.1) to filter the high-frequency signals (i.e. tidal and inertial motions) and focus on correcting the subtidal processes. Notice that tides in this area have very low amplitudes of the order of a few centimeters only. Daily means of radial and total currents are computed independently for each data type from the hourly observations. For the total currents, the daily mean is only considered at grid points where, during each day, at least 50% of hourly measurements are both available and flagged as good, as also used by Lorente et al. (2015). In the case of the radials, we considered a threshold of 25% for computing the daily mean.

As stated above, at each grid point, the hourly total currents are calculated using all available radial observations within a radius of 6 km. Consequently, some total observations could be computed using different radial grid points within this radius for each hour that individually do not satisfy the threshold of 50% imposed for the total velocities. Therefore, using the same threshold to calculate the daily means of both observations could lead to patches with available reconstructed daily mean total currents but no daily mean radials available. This is why we decided to use a less restrictive threshold for radials to have better spatial coverage, consistent with that of the total observations in the area covered by both antennas.

3.2.2 Regional model configuration

The Western Mediterranean OPerational system (WMOP, Juza et al. (2016); Mourre et al. (2018)) is a high-resolution regional configuration of the ROMS (Regional Ocean Modelling System) model (Shchepetkin and McWilliams, 2005) for the western Mediterranean Sea. The spatial coverage spans from Gibraltar strait on the West to the Sardinia Channel on the East (6ºW-9ºE, 35ºN-44.5ºN, see Fig. 3.2) with a horizontal resolution around 2 km and 32 vertical sigma levels (resulting in a vertical resolution between

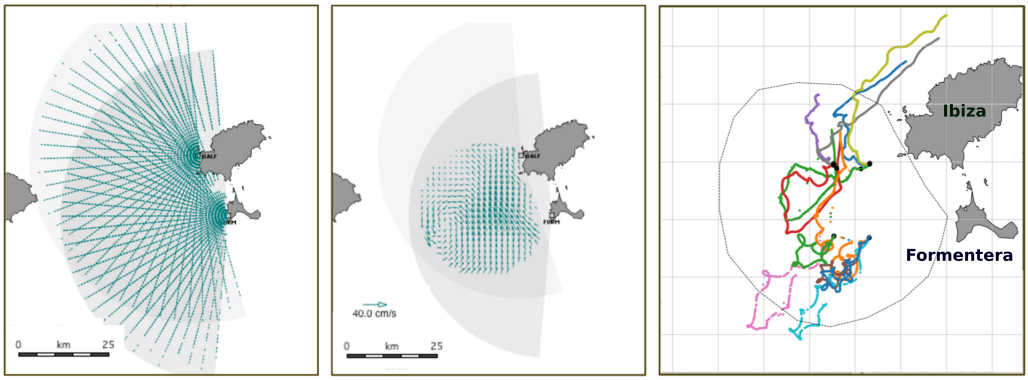

Figure 3.1 *Map of the Ibiza Channel showing the HFR coverage area for radial (left panel) and total (central panel) currents, together with the position of the two antennas (GALF and FORM). The right panel shows the 13 drifters used for validation and their trajectories within the first 6 days after deployment. Each drifter has a randomly assigned color. Dots indicate start locations of the trajectories.*

1 and 2m at the surface). The WMOP system is used to produce daily forecasts of the regional ocean circulation, which is used for a wide range of applications including search-and-rescue and analysis of plastic, parasite or larval dispersion (Calò et al., 2018; Ruiz-Orejón et al., 2019; Cabanellas-Reboredo et al., 2019; Compa et al., 2020; Torrado et al., 2021; Kersting et al., 2020; Révelard et al., 2021).

The vertical mixing coefficients are set using the Generic Length Scale (GLS) turbulence closure scheme (Umlauf and Burchard (2003), with parameters p=2.0; m=1.0; n=-0.67 as in line 1 of their Table 7). The bathymetry is derived from a 1' database (Smith and Sandwell, 1997). The simulation used in this study is initialized from and nested within the larger scale Copernicus Forecasting System (CMEMS MED-MFC), with a 1/16º horizontal resolution (Simoncelli S., 2017). The atmospheric forcing is provided every 3 hours at $1/20°$ resolution by the Spanish Meteorological Agency (AEMET) through the HIRLAM model (Undén et al., 2002). These fields are used to compute surface turbulent and momentum fluxes through bulk formulae. Atmospheric pressure forcing is neglected to avoid SSH high-frequency variability issues. Inflows from the six major rivers in the region are considered as point sources, using daily climatological values. Tides are not considered in the model.

A multi-year free-run hindcast spanning the period from 2009 to 2018 (Mourre et al., 2018; Aguiar et al., 2020) has been used as control simulation. This simulation also provides the initial state for the data assimilative simulations starting on 20 September 2014. Fig 3.2 shows the mean surface field of the control run during the simulation period (20 September to 20 October 2014) together with the mean surface currents measured by the HFR for the same period. The HFR observations depict an average southward current west of 0.8E. This current is deviated towards the south-east of 38.7N, and the flow is directed northward in the eastern side of the coverage area, close to Ibiza and Formentera coast. The control run represents this overall pattern, but with a significant overestimation of the mean velocities and a spatial mismatch of the eastward deviation of the flow (this deviation occurs too much to the east in the model).

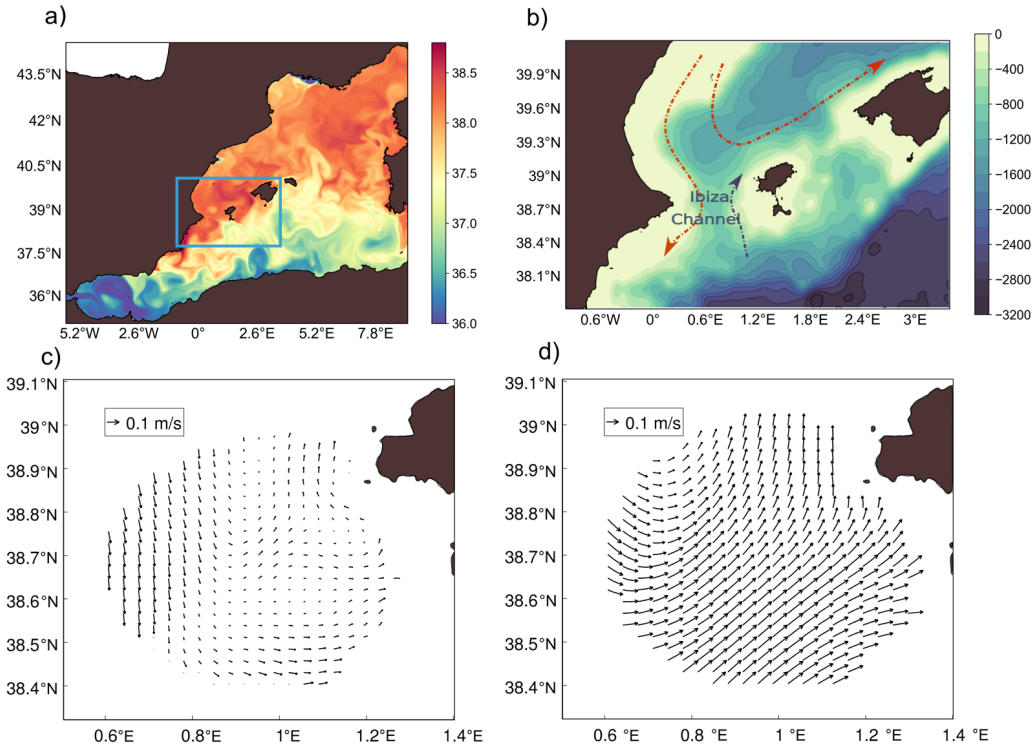

Figure 3.2 *Illustration of the modelling domain and study area. a) WMOP sea surface salinity (6 October 2014). The Ibiza and Mallorca Channel area is delimited by the blue rectangle. b) Bathymetry (m) and main circulation features in the Ibiza Channel. c) Mean HFR surface currents over the whole simulation period (20 September to 20 October 2014). d) Mean surface currents over the whole simulation period computed from the model.*

3.2.3 Data Assimilation System

The assimilation scheme employed here is the local multi-model Ensemble Optimal Interpolation (EnOI) employed in Hernandez-Lasheras and Mourre (2018). It is a form of the EnOI, which has been a widely used scheme since it represents a cost-effective alternative compared with more complex methods as the Ensemble Kalman Filter or the 4Dvar (Oke et al., 2002; Evensen, 2003; Counillon and Bertino, 2009). EnOI is a 3D sequential assimilation method that allows the use of a large ensemble size together with localization. A stationary ensemble of model simulations is used to estimate background error covariances. The WMOP-DA system consists of a sequence of analyses (model updates given a set of observations) and model forward simulations.

For each analysis, the state vector $\mathbf{x} = (\mathbf{T}_{i,j,k}, \mathbf{S}_{i,j,k}, \mathbf{u}_{i,j,k}, \mathbf{v}_{i,j,k}, \mathbf{SSH}_{i,j})^{\mathbf{T}}$, contains the model trajectory, i.e., the prognostic model variables at all wet gridpoints i, j, k. During the analysis step, the state vector $\mathbf{x}^a$ is updated according to Eq. (3.1), where $\mathbf{x}^f$ is the background model state vector, $\mathbf{H}$ is the linear observation operator projecting the model state onto the observation space and $\tilde{\mathbf{K}}$ is the Kalman gain estimated from the sample covariances (Eq. 3.2). $\mathbf{y}$ is the vector of observations. Matrices $\tilde{\mathbf{P}}^f$ and $\mathbf{R}$ are the error covariance matrices of the model and the observations, respectively.

$$\mathbf{x}^a = \mathbf{x}^f + \tilde{\mathrm{K}}(\mathbf{y} - \mathbf{H}\mathbf{x}^f), \tag{3.1}$$

$$\tilde{\mathbf{K}} = \tilde{\mathbf{P}}^f \mathbf{H}^T (\mathbf{H}\tilde{\mathbf{P}}^f \mathbf{H}^T + \mathbf{R})^{-1}, \tag{3.2}$$

$\tilde{\mathbf{P}}^f$ contains the background error covariances (BECs). In our approach we estimated the BECs by sampling three long-run simulations of the WMOP with different initial and boundary forcing provided by the CMEMS MED-MFC (Simoncelli S., 2017) and CMEMS GLO-MFC (Lellouche et al., 2018) forecasting systems and varying momentum and diffusion parameters. An ensemble of 80 realizations is considered in each analysis. Each ensemble member is randomly extracted from the three different long-run simulations within a temporal window of 90 days centered on the day of the analysis for the different years covered by the three long-run simulations. The seasonal cycle is removed from the multivariate fields before computing the ensemble anomalies to limit the effects of large scale correlations, mainly in terms of surface temperature. This way, we obtain multivariate, inhomogeneous and anisotropic 3-dimensional model BECs characteristic of the mesoscale variability. We used a domain localization of 200 km, corresponding to the average distance between two Argo profiles in the Western Mediterranean Sea. An independent analysis is performed for each water column of the model domain, considering only the observations within the localization radius.

We used a diagonal observation covariance error matrix **R**. The observation error for all HFR observations has been considered the same, with a value of $0.1\mathrm{m/s}$ (accounting for instrumental and representativity errors). This value is consistent with local comparisons against surface currents measurements from a point-wise currentmeter (1.5 m depth) and a downward-looking ADCP (first bin at 5 m depth) carried out by Lana et al. (2016), which reported a RMSD between 0.07 and 0.12 m/s. This value of 0.1m/s was fixed in our experiments as it yielded to a proper correction of surface currents, without degrading the vertical structure. Once the observation error was set for HFR total observations, we performed new experiments to evaluate the potential differences between the total and radial observations. Total observations were interpolated and projected to generate synthetic radial observations containing the exact same information as the totals but with a radial-like pattern in the area covered by both antennas. The assimilation of these total and synthetic radial observations using the same observation error led to almost identical results in surface fields and vertical structure, with complex correlation of 0.92 and a RMSD of 0.02 m/s obtained between both analysis fields in the HFR grid points. Based on these results we decided to use the same observation error for both types of observations.

The state vector equivalents of HFR radials are obtained using the following equation:

$$\mathbf{H}\mathbf{x}^f = \mathbf{u}_\mathbf{x} \cos\boldsymbol{\alpha} + \mathbf{u}_\mathbf{y} \sin\boldsymbol{\alpha}, \tag{3.3}$$

where u_x and u_y are the model surface velocity components interpolated at the observation point, and α denotes the angle (anti-clockwise towards the east) pointing from an antenna station to a certain location.

A 3-day assimilation cycle is applied with different time windows for each source of observation as explained in the following section. In each analysis (*day n*) the daily average field is employed as background and two different initialization approaches (Fig. 3.3) have been applied to restart the model after the analysis. Sequential assimilation methods are affected by initialization issues, as primitive equation models are sensitive to discontinuous changes in their model fields (Oke et al., 2002). These discontinuities may introduce artificial waves or structures in the model that affect the quality of predictions. Different strategies have been proposed to address this problem (Sandery

et al., 2011; Yan et al., 2014).

In the first approach, the simulation for *day n+1* restarts directly from the results of the analysis. The second approach, which will be referred to as *nudging* consists in running again the *day n* applying a very strong nudging (time scale of one day) towards the temperature, salinity and SSH fields provided by the analysis. Notice that the nudging is not applied to the velocity fields. These are adjusted by the model itself according to its dynamics. This procedure reduces the model corrections but guarantees updated multivariate fields closer to model equation balances, which limits instabilities. This set-up is very similar to the one employed by Oke et al. (2007) in the Bluelink system.

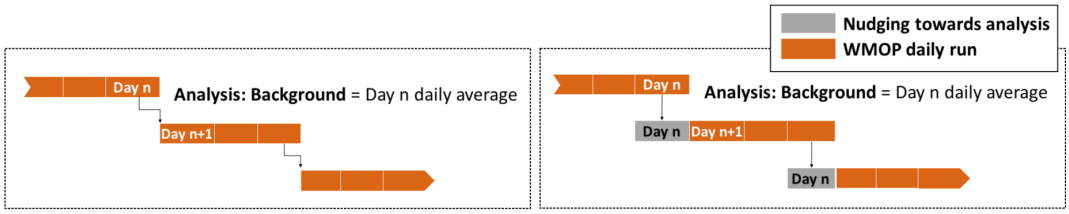

Figure 3.3 *Data Assimilation procedure, illustrating the two initialization methods and the 3-day cycles. The diagram on the left describes the direct initialization strategy from the analysis. The diagram on the right describes the Nudging strategy for initialization. Orange rectangles represent each 1-day run of WMOP. Grey rectangles represent the 1-day run of WMOP in which a strong nudging towards the results of the analysis is applied.*

3.2.4 Simulations

Seven simulations of WMOP are used to investigate the impact of both HFR observations and initialization methods (Table 3.1). The period selected for the simulation experiments covers one month, from September 20^{th} to October 20^{th} 2014, assimilating different sets of observations every 3 days. During this period a total of 13 satellite-tracked surface drifters (Tintoré et al., 2014) were deployed in the area covered by the HFR and used as independent data for validating the numerical experiments (Fig. 3.1). We adopted the operational prediction setup of WMOP, considering only observations before the analysis date. Notice that a "retrospective analysis" framework considering a time window centered on the analysis date could slightly improve the results presented in this chapter. However, since our objective is to implement this method for daily predictions, the operational setup has been selected. Satellite SLA (sea level anomalies), SST (sea surface temperature) and T-S (temperature and salinity) Argo profiles, defined as the Generic Observing sources (GO), are assimilated in all these simulations. The SLA consists in along-track L3 multi-satellite reprocessed observations provided by CMEMS. We consider a 3-day window for SLA observations. The SST comes from a L4-GHRSST foundation SST product distributed by JPL-MUR (NASA/JPL, 2015). The foundation SST is the temperature free of diurnal temperature variability, corresponding to the temperature of the surface just before the daily heating by the sun. Since the model daily average contains the signature of the diurnal cycle, this effect needs to be accounted for in the representativity error. This is approximated by computing the variance of the difference between the model SST field at 8 a.m. and the

daily average field used as background for each of the grid points. The ultra-high 1 km resolution gridded fields have been smoothed and interpolated to a 10 km grid to limit the number of observations, while still representing the effective scale that this SST product can resolve (Chin et al., 2017). For the T-S Argo profiles we have considered a 5-day time-window, which corresponds to the nominal time of Mediterranean Argo floats cycles. For each profile, values are binned vertically to obtain a single value for each model grid cell. The variance of the data within a bin is used as the vertical representation error, which is added to the horizontal one, assumed to be $0.25°C^2$ and 0.05^2 for temperature and salinity measurements, respectively.

A control run (CR) without data assimilation has been used as benchmark to assess the performance of the different assimilation experiments. We called GNR the simulation in which we only assimilated GO. Additionally, four other simulations assimilating HFR data together with GO have been generated. In all four cases we assimilate daily averages to remove the impact of inertial oscillations and tides, which are not the focus of this study. Daily averaged fields from the model are used as background for the analysis. TOT simulation employs HFR totals, computed as described before. We called RAD the simulation assimilating all possible daily mean radial observations.

Data assimilation experiments have been repeated using both types of initialization for every dataset. Our analysis will first evaluate the impact and trade-offs of the different kind of HFR observations when using the direct restart from the analysis procedure. Then, the impact of the nudging initialization method will be specifically discussed.

Experiment	Assimilated observations
CR	None
GNR	SLA, SST, TS
TOT	SLA, SST, TS, HFR totals
RAD	SLA, SST, TS, HFR radials

Table 3.1 *Basic description of the experiments, indicating the dataset used in the simulations.*

3.3 Results

3.3.1 Assessment of the impact of DA on SST, SLA and T-S profiles over the whole domain

To evaluate the general performance of the DA, all one-month-long experiments are first compared against SLA, SST and Argo TS profiles over the whole modelling domain. For each experiment, the WMOP fields are interpolated onto the observation points. Each day, the model SLA has been assessed against along-track daily multi-satellite altimetry observations provided by CMEMS. Model daily mean fields at the observation points are considered. The satellite SST product is compared to the model SST at 8 a.m. to reduce the potential impact of the diurnal cycle. For the comparison of the model and the Argo T-S profiles, the available daily observations are compared against model daily mean fields. The closest grid point of the model has been considered. Due to the backward-in-time assimilation window, the observations used for the validation have not been previously assimilated. However, they can not be considered

as fully independent since the data employed for the validation come from the same platforms that provide the assimilated measurements.

Taylor diagrams (Taylor, 2001) are presented here for the evaluation of the simulations. They illustrate the correspondence between model and observations in terms of correlation coefficient, centered root mean square difference (CRMSD) and standard deviation. However, note that the diagram does not represent the mean error between the observations and the model, which has been examined separately. The magnitude of the SST mean error decreases from -0.29°C to -0.14°C , representing in all simulations less than the 14% of the total RMSD [1]. The mean error between the CR and the Argo profiles is 0.4°C and -0.13 for temperature and salinity respectively, representing less than 8% of the RMSD in both cases.

The use of DA results in a significant improvement of both the SLA and SST fields, as shown in Fig. 3.4. For both data sources the symbols corresponding to each simulation assimilating data overlap, meaning that the validation metrics are very similar for all of them. For the SLA it leads to a significant increase in the correlation, with values from 0.42 to around 0.70, and a 30% reduction in the CRMSD for all the experiments with DA. Notice that the model SLA presents a relatively large mean error, with a value of around -0.07m. Discrepancies are common when comparing models to altimetry due to differences in the mean sea level. This mean error, which persists after DA, accounts for the difference between the mean dynamic topography of the model and observations. This way, the reduction in the RMSD is mostly due to reductions of the CRMSD, which can be observed in the diagrams. Concerning the SST, we obtain a similar error reduction in terms of centered RMSD, of the order of 30% closer to observations when using DA. An increase in correlation is also obtained, from 0.82 to around 0.92 when compared with the CR. We do not observe a significant difference between the simulations using different datasets.

Similar conclusions are obtained when examining the Taylor diagrams focusing on Argo temperature and salinity profiles (Fig. 3.5). Although the CR simulation shows a very high correlation with observations (0.88 and 0.95 for temperature and salinity respectively), this correlation is further increased for the experiments with DA. A CRMSD reduction of more than 35% is obtained for both salinity and temperature observations in all data assimilative experiments. The diagram for the salinity shows a decrease in the standard deviation with DA and slight differences between RAD and the other two simulations.

The impact of the assimilation on the different fields has been also evaluated considering only observations surrounding the IC area, leading to similar results.

3.3.2 Eulerian assessment of the impact of DA on surface currents

To evaluate the DA capabilities to improve the representation of surface currents, we performed an Eulerian analysis in the HFR coverage area. WMOP surface daily mean velocities are compared against HFR totals daily mean fields. The total observations are derived from the radial data, as described in section 3.2.1. We compute the daily mean field only at those points that provide more than 50% of hourly data and then interpolate the model to HFR observation grid points. As for the SLA, SST and Argo TS profiles, the validation can not be considered fully independent since we use the

[1] The RMSD has a contribution from the bias and CRMSD according to the following formula: $RMSD^2 = CRMSD^2 + bias^2$

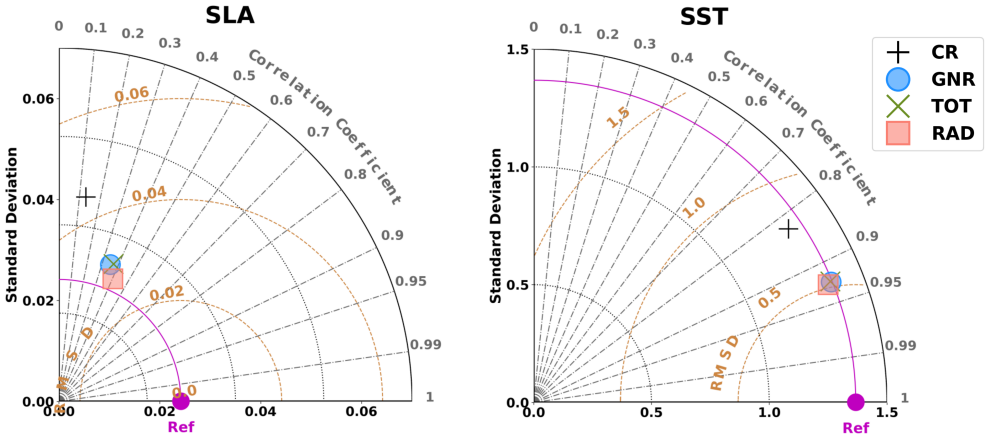

Figure 3.4 *Taylor diagrams comparing models and observations in terms of SLA (left) and SST (right) over the whole modelling domain. X and Y axis represent the standard deviations of the data. Distance from the reference point located on the X axis (noted as Ref. in magenta) represents the centered root mean square deviation (CRMSD). Correlation between observations and model increases clockwise. Symbols represent the different simulations, as specified in the legend*

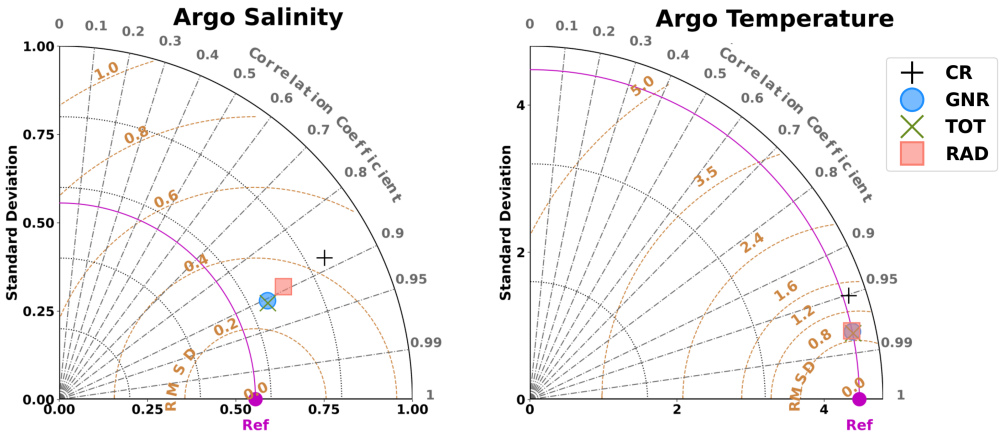

Figure 3.5 *Same as Figure 3.4 for Argo temperature (left) and salinity (right) profiles.*

same observing platform. However, the data used for validation at a given time have not yet been assimilated in the model.

We first analyze the performance in terms of surface currents by using the Taylor diagrams for the velocity components (Fig.3.6). The zonal velocity component experiences a strong correction with the assimilation of GO. Specifically, the CRMSD is reduced by 28% while the correlation increases from 0.28 to 0.44. This performance is further improved by the two experiments using HFR data, with more than a 40% reduction in CRMSD. While TOT experiment exhibits the largest error reduction, RAD provides the best correlation with observations (0.7), compared to 0.63 obtained by TOT..

Considering the meridional velocity component GNR has a lower correlation and higher CRMSD than the CR. Here, the use of HFR observations is necessary to reduce the difference between model and observations. The correlation slightly increases with

the assimilation of HFR, with the best results obtained for TOT (0.47). Moreover, the standard deviation and CRMSD displays a significant reduction (27% for TOT and 19% for RAD).

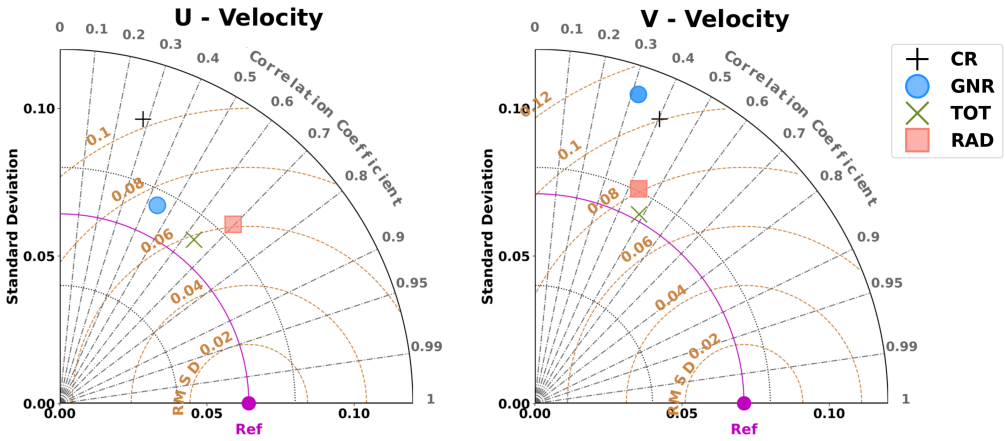

Figure 3.6 *Taylor diagrams for WMOP simulations compared to HFR surface currents observations. Separate diagrams for each velocity component: U (left) and V (right). The symbols represent the different simulations, as specified in the legend.*

Figure 3.7 shows the spatial distribution of the surface current speed mean error (bias), defined as the difference between the HFR and the model in terms of the module of the velocities at each grid point. Positive error reflects that observation mean values are larger than model estimates. The control run, CR, overestimates the currents in most of the domain with the exception of a small area on the western side. The mean bias over the whole HFR domain is of 0.10 m/s (see Table 3.2). DA corrects the bias in all 3 experiments. The assimilation of GO leads to a reduction of the error over the whole domain, with a mean value of 0.05 m/s. A further reduction is achieved when assimilating HFR velocities. RAD has a mean bias of 0.03 m/s, which is particularly higher near the Ibiza antenna, while TOT has the lowest bias, with a mean value of 0.01 m/s.

Experiment	**Bias (m/s)**	**nRMSD**	**RMSD (m/s)**	**U-RMSD (m/s)**	**V-RMSD (m/s)**
CR	0.10	1.00	0.21	0.13	0.16
GNR	0.05	0.79	0.17	0.08	0.14
TOT	0.01	0.52	0.11	0.06	0.09
RAD	0.03	0.57	0.12	0.06	0.10

Table 3.2 *Bias, normalized RMSD and total RMSD between model and HFR surface currents speed. The two columns on the right correspond to the RMSD for the zonal and meridional components.*

3.3.3 Lagrangian assessment of the impact of DA on surface currents

As previously stated, 13 surface drifters were deployed in the HFR coverage area during the simulated period, as described in Lana et al. (2016). Three different kinds

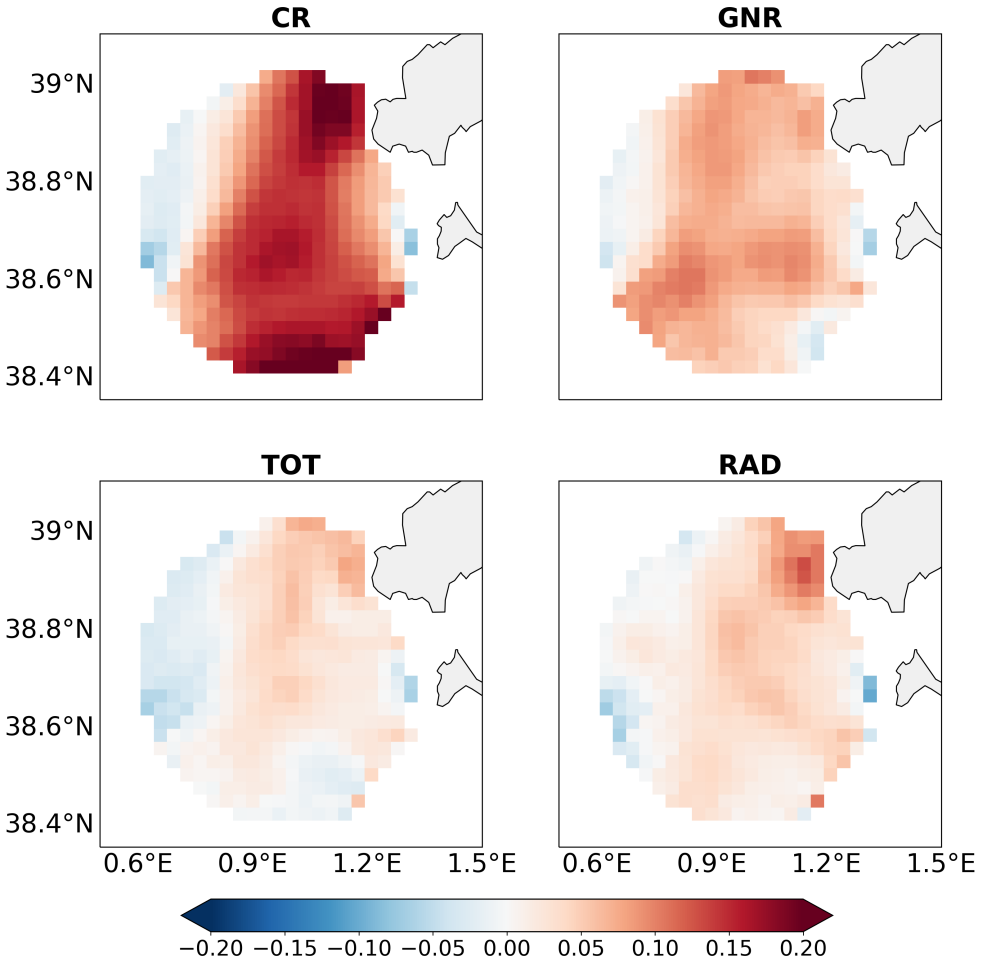

Figure 3.7 *Mean total current speed error (bias, m/s). The mean model speed is subtracted from observations at each grid point. Positive values indicate that the model overestimates the observations in average.*

of surface drifters (ODi, MDOi and CODEi) were employed, all drifting at a depth between the surface and 1m. No significant wind drag is expected for these drifters (more details can be found in Révelard et al. (2021) or Barth et al. (2021)). Virtual particle drifts were then computed using model surface currents. For each experiment, and for eight consecutive days (from October 1^{st} to October 8^{th}), 1000 neutrally buoyant particles were launched at each of the positions of the 13 drifters at 00:00 each day. Lagrangian tracks were simulated using Ocean Parcels (Lange and Van Sebille, 2017) and 5 days period of WMOP velocity fields (at 3-hours resolution). Additionally, we included a diffusion term using a Brownian motion scheme with the objective of representing the impact of the subgrid processes not resolved by the model. After each advection step the diffusion is imposed using a random distribution with a diffusion coefficient of $50\ \mathrm{m}^2/\mathrm{s}$, in line with recent Lagrangian studies using the WMOP model (Cabanellas-Reboredo et al., 2019; Compa et al., 2020; Kersting et al., 2020; Torrado et al., 2021). We have verified that our results are not significantly affected by this value of the diffusion coefficient, which has a significant impact on the spread of the

trajectories but not on the path of the mean trajectory.

The distance between real drifters and the center of mass of each set of the 1000 Lagrangian particles is computed at each time step and a skill score (Eq. 3.4 and 3.5) is given for each drifter every day following the description made by Liu and Weisberg (2011). A non-dimensional index s is calculated based in the normalize cumulative Lagrangian separation distance, from purely Lagrangian parameters (Eq. 3.4), where d_i is the separation distance between the modeled and observed endpoints of the Lagrangian trajectory at time step i after initialization, l_{oi} is the length of the observed trajectory, and N is the total number of time steps.

$$s = \sum_{i=1}^{N} d_i \Big/ \sum_{i=1}^{N} l_{oi} \tag{3.4}$$

$$SS = 1 - s \tag{3.5}$$

Trajectories are also simulated using the hourly DIVAnd reconstructed velocity fields presented in Barth et al. (2021). DIVAnd is a n-dimensional variational analysis method which is used here to reconstruct hourly 2D vectorial fields from radial observations. It was shown to improve the reconstruction compared to the Open-boundary Modal Analysis (Kaplan and Lekien, 2007). Figure 3.8 provides a few examples illustrating the Lagrangian prediction capacity for the different simulations. Each panel shows the trajectories of the drifter and the center of mass of the virtual particles for each experiment for 48 hours of simulation. These plots illustrate the diversity of situations associated with the spatio-temporal variability of the surface ocean velocities. In particular for panels b), c), d), and e), TOT displays a very good agreement with the observations, resulting in the best performance over all the simulations. However, it is worth noting that this behavior is not systematic and the simulations assimilating HFR sometimes fail in providing the best trajectories (panels a) and f)).

Figure 3.9 shows the skill score for all experiments (4 model simulations + DIVAnd fields) for a forecast horizon of 48 h. Each point is located at the initial position of the particles at the beginning of the Lagrangian simulations and represents the value of the skill score of the center of mass of the cloud of virtual particles. For trajectories with a cumulative separation distance larger than the cumulative distance traveled by the particles, the model has a negative skill score. On the other hand, values close to 1 indicate a nearly perfect match between the drifter and virtual trajectories. Values of the mean skill score are given in table 3.3, where the mean skill score is computed separately for (i) the whole trajectories; (ii) the trajectories starting inside the HFR coverage area and (iii) trajectories whose initial position is outside the HFR area. The top left panel of Figure 3.9 shows the spatial distribution of the SS for the CR, highlighting multiple trajectories inside the HFR coverage area for which the model has no skill ($SS < 0$) according to Liu and Weisberg (2011) criteria, resulting in a negative mean value (-0.35), and trajectories outside for which the model has some skill, with scores over 0.5, resulting in a mean values of 0.36. All experiments present a larger SS outside the HFR coverage area. This in mainly due to the characteristics of the trajectories north of the island of Ibiza, where the circulation is dominated by the Balearic current, with a more steady northeastward flow generally better reproduced in the model, as previously described by (Révelard et al., 2021).

In general, all experiments with DA improve the trajectories. In particular, GNR

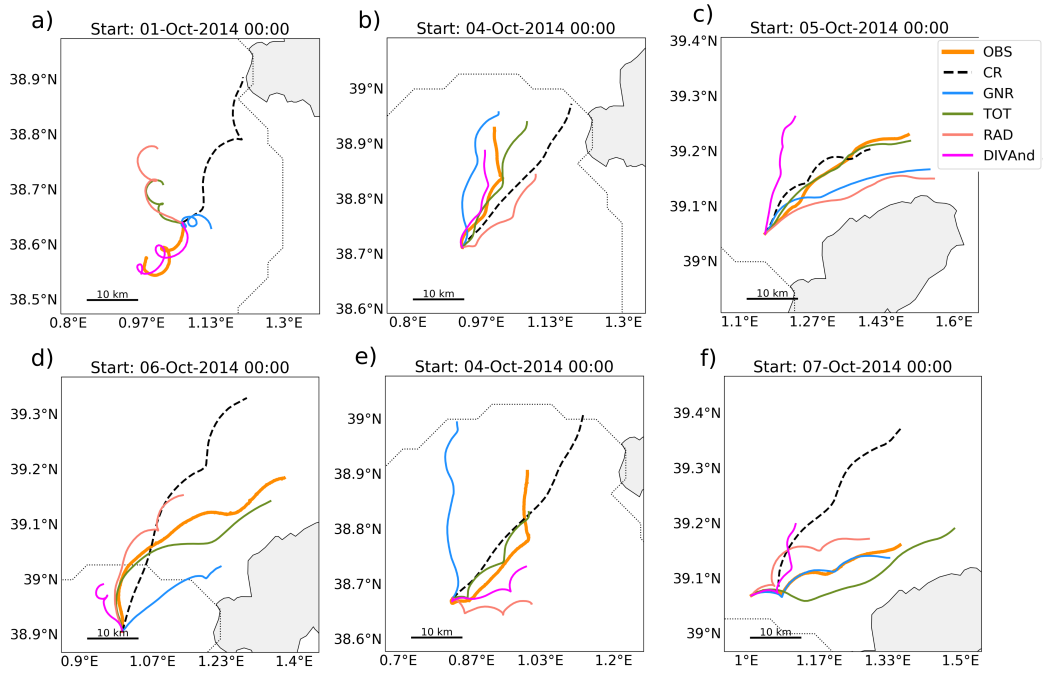

Figure 3.8 *Map showing two-day satellite-tracked drifters and model trajectories derived from different DA experiments for different dates. Model trajectories represent the trajectories of the center of mass of the 1000 particles launched at the drifter position at 00:00 of the indicated starting day.*

Experiment	**SS Whole domain**	**SS Inside**	**SS Outside**
CR	-0.16	-0.35	0.36
DIVAnd	0.45	0.51	0.28
GNR	0.19	0.05	0.58
TOT	0.45	0.41	0.57
RAD	0.42	0.36	0.59

Table 3.3 *Average Skill Score for the different experiments, over the whole domain, and inside and outside the HFR coverage area.*

increases the skill score compared to CR with values increasing from -0.16 to 0.19 over the whole domain. Note, however, that inside the coverage area not all trajectories are properly represented by the model (Fig. 3.9). The assimilation of HFR data along with GO further increases the skill score. The improvement is particularly significant inside the HFR domain, where most of the trajectories have positive SS. TOT has the best results among the model experiments, with a mean value of 0.41 inside the coverage area, which is better than RAD (0.36). Outside the coverage area all data-assimilative simulations lead to a similar performance (SS around 0.58), representing an improvement with respect to CR (SS=0.36). Assimilating HFR data does not lead to any significant improvement nor degradation of the model performance outside this coverage area with respect to GNR.

The average separation distance is computed according to Equation 3.6, where $n_{drif} = 13$ is the number of drifters, $n_{part} = 1000$ is the number of particles and $\mathbf{x}^{\mathrm{d}}$ and $\mathbf{x}^{\mathrm{v}}$ are the positions of the real drifter and the corresponding virtual particle

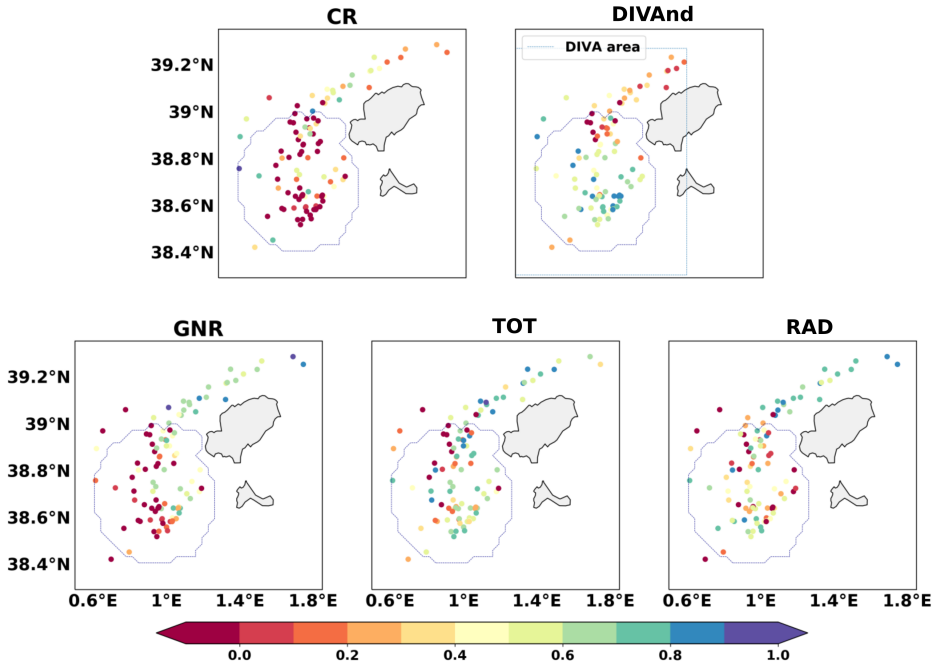

Figure 3.9 *Scattered dots represent the skill score (Liu and Weisberg, 2011) of each simulation to represent a drifter trajectory. The dot position is the starting point of each Lagrangian simulation. Values lower than 0 mean the simulation has no skill at representing that specific drifter trajectory according to the metric used, while values close to 1 mean a perfect performance of the model. Skill score has been calculated for 48 hours.*

respectively (Figure 3.10). For each 5 day trajectory, the mean separation distance is first computed averaging over the number of drifters, providing a single distance as a function of time $d(t)$ for the 13 drifters (Eq. 3.6). Then, the four values of $d(t)$, one for each of the four simulations starting in consecutive days are averaged.

$$d(t) = \frac{1}{n_{drif}} \sum_{i=1}^{n_{drif}} \left(\left| \mathbf{x_i}^{\mathbf{d}}(\mathbf{t}) - \frac{1}{n_{part}} \sum_{j=1}^{n_{part}} \mathbf{x_{ij}}^{\mathbf{v}}(\mathbf{t}) \right| \right) \tag{3.6}$$

The mean distance between virtual and real drifters is significantly reduced when DA is applied. The assimilation of GO efficiently helps to reduce the mean separation distance, with a reduction of 31% after 48 hours compared to CR (18.9 versus 27.2 km). Consistent with the previous analysis, the assimilation of HFR total observations along with the GO further increases the performance, leading to the lowest mean separation distance (12.8 km), with a 53% reduction compared to the CR. The use of radial observations also leads to a high reduction of the mean separation distance (48%), which is reduced to 14.3 km after 48hr.

DIVAnd simulations present a mean distance of 8.4 and 17.3 km after 24 and 48 h respectively, affected by a significant number of trajectories outside of the HFR coverage area and so, not properly constrained by the reconstruction algorithm.

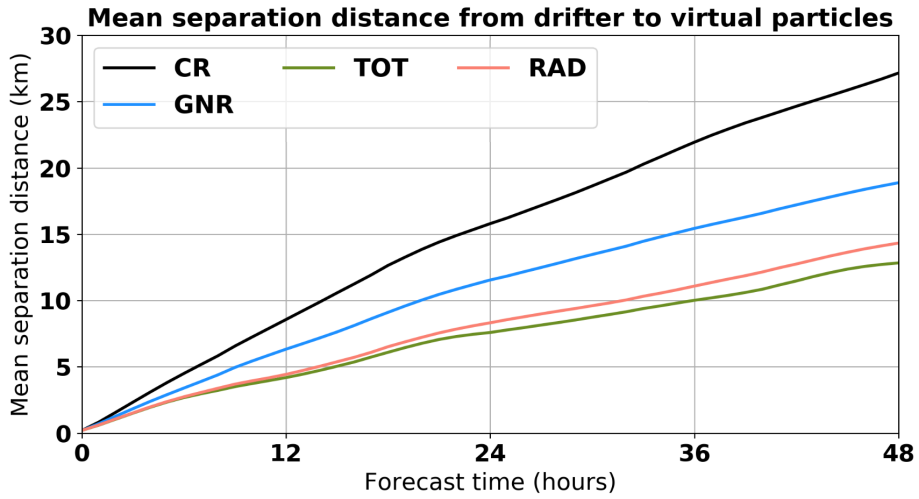

Figure 3.10 *Mean separation distance between drifter and the center of mass of virtual particles using the direct restart from analysis*

Experiment	**Bias (cm/s)**	**nRMSD**	**RMSD**	**U-RMSD**	**V-RMSD**
CR	10.42	1.00	0.21	0.13	0.16
GNR-N	5.04	0.75	0.16	0.09	0.12
TOT-N	2.21	0.59	0.12	0.07	0.09
RAD-N	2.16	0.58	0.12	0.07	0.10

Table 3.4 *Same as Table 3.2 for the simulations applying a nudging towards the analysis.*

3.3.4 Impact of the nudging restart strategy

Overall, the results in the whole domain comparing to satellite and Argo observations are similar to those obtained for the simulations restarting directly from the analysis. The improvement is slightly lower due to the nudging step, but all data assimilative simulations provide comparable metrics. The reduction of the RMSD compared to the CR is around 8% for the SLA, while for the SST is reduced around 30%. Considering Argo profiles, the reduction of the RMSD is of 35% for all simulations, both for temperature and salinity.

Table 3.4 presents the bias and RMSD for the model surface current speed and the zonal and meridional components. This has to be compared with Table 3.2, which shows the results for the simulations restarting from the analysis. We can observe a slight improvement for the $GNR-N$ simulation when using the nudging initialization in comparison to restarting directly from the analysis, with a reduction of both the bias and the RMSD. While this initialization method also helps to reduce the bias compared to direct restart from the analysis for RAD , this is not the case when using total observations.

The Lagrangian assessment confirms these results, reflecting the usefulness of HFR data to correct surface currents using this initialization method even when the nudging is only performed towards the SSH and TS fields. The SS for the $GNR-N$ simulation (Table 3.5) increases significantly inside the coverage area while decreasing outside, with an average value of 0.39, larger than the value of 0.34 obtained with the other

approach. The correction obtained using HFR Total velocities together with GO is slightly degraded using the nudging approach both inside and outside the coverage area, with an average SS of 0.41 compared to 0.45 with the direct restart from analysis. On the other hand, RAD has a better SS when using the nudging approach. The average SS inside the radar domain increases from 0.36 to 0.38, while outside the domain it slightly decreases from 0.59 to 0.57.

Experiment	**SS Whole domain**	**SS Inside**	**SS Outside**
CR	-0.16	-0.35	0.36
GNR-N	0.28	0.18	0.54
TOT-N	0.41	0.35	0.56
RAD-N	0.43	0.38	0.57

Table 3.5 *Same as Table 3.3 for the simulations applying a nudging towards the analysis.*

The mean separation distance after 48h when assimilating GO is also reduced from 18.9 to 16.7 km when using the nudging initialization. Although the assimilation of total HFR velocities further decreases the mean separation distance, results are slightly degraded when using the nudging compared to restarting directly from the analysis, with a mean distance of 14.0 km after 48 h. On the contrary, the assimilation of HFR radials benefits from the nudging approach, reducing the mean separation distance from 14.3 to 13.4 km after the first two days compared to the direct restart from analysis (which represents a 51% reduction in comparison to CR) and giving the best results among all simulations using this initialization method.

3.4 Discussion

The assimilation of high-resolution HFR surface currents observations in a reduced part of the modelling domain could have a negative effect on the rest of the variables under the effect of spurious model error correlations. While in Stanev et al. (2015, 2016) the positive outcome of the data assimilation extends beyond the HFR covered area, Zhang et al. (2010) showed that the assimilation of HFR led to an improvement of surface currents, but with a degradation of the sub-surface temperature forecasts, in their experiments. Sperrevik and Christensen (2015) evidenced that using TS profiles along with HFR observations led to better results, as they control the density fields while adding a constraint on the circulation. Here we show that the assimilation of local HFR (both totals and radials) observations along with the generic ones does not degrade the improvement on SLA, SST fields and on Argo TS profiles achieved over the whole domain when assimilating only the generic observations. The results obtained for all experiments with DA show similar performance in this sense. Differences mostly depend on the type of initialization employed. Nevertheless, this work is mainly focused on the study of surface currents and thus, the impact on sub-surface fields has not been deeply analyzed. CTD casts or glider data in the region should help to complete the assessment in future studies.

We have used DIVAnd reconstructed fields as a benchmark for our Lagrangian validation. These hourly fields properly represent the inertial oscillations, compared to other gap-filling techniques (Barth et al., 2021), and we consider it as the best possible

high resolution representation of the surface currents in the area which allows the simulation of Lagrangian trajectories. It is very positive that the skill scores obtained for the HFR DA experiments are very close to that obtained by DIVAnd. While DIVAnd outperforms the capabilities of the WMOP DA system inside the coverage of both HFR antennas, it is the opposite outside this region, which demonstrates the capacity of the model to improve the represention of the currents beyond the HFR coverage area. The assimilation of GO, in particular SLA, constrains the geostrophic circulation, leading to a better representation of the Balearic current and an increase of the SS in that area. The importance of this constraint is highlighted when comparing with DIVAnd-derived trajectories, which do not properly represent these features. While the mean SS for the DIVAnd-derived trajectories inside the area is 0.53, it drops to 0.29 outside of it, being significantly lower than all model-derived trajectories. This behaviour is consistent with Barth et al. (2021) results, which show that the DIVAnd reconstructed fields outside the area covered by both HFR antennas are much less reliable. Our results demonstrate the utility of dynamical models assimilating high-resolution observations as good alternatives to data-driven short-term forecasting methods, due to their capacity to extend the correction beyond the observation coverage area. They also show the importance of combining HFR and altimeters observations which help to constrain the geostrophic circulation over a wider area.

Two different initialization strategies have been evaluated. While restarting directly from the analysis may introduce some high-frequency and spurious waves or instabilities in the system due to inconsistencies between the corrected fields and the model equations, it considers an initial state which is closer to observations. On the other hand, the nudging strategy provides a more conservative framework, in which the model dynamics are better respected but with the drawback that some of the correction achieved with the observations may be lost. Overall, both approaches show similar results leading to a reduction of the RMSD over the whole domain. As in the case of the direct restart from the analysis, the use of the $nudging$ strategy also leads to an improvement of the predictions of surface currents when assimilating HFR observations, compared to the simulation that only uses generic data sources. It is important to point out that, in our case, nudging is only applied towards the temperature, salinity and sea surface height fields, but not towards the velocity fields, to avoid model instabilities. Therefore, the assimilation of the surface currents enables to correct the density fields, which in turn improves the surface velocities due to the model initial adjustments.

The $nudging$ strategy limits the possible shocks and anomalous gradients that may be generated in the analysis, so that the solution remains closer to the physical balances. We found that it was not optimal for surface currents prediction when using HFR total velocities but a better choice for radial data. This is probably due to the fact that reconstructed total velocities are already smoothed out through a pre-processing step contrarily to the case of radial data, which are more noisy and then directly benefit from the smoothing effect of the nudging approach. The nudging strategy appears to be a good solution for operational purposes, when the occurrence of noisy data tends to be more frequent. It may also be a good choice for systems depending on operational data sources for which HFR antennas, for instance, may not work during certain periods or satellite and Argo data may not be available on time. It could also be less sensitive to potential errors in data in cases where near real-time observations could be affected by significant errors.

The observation error was considered the same for total and radial currents in this

study, as explained in section 3.2.3, without considering any spatial variability. Some authors used spatially variable observation errors depending on whether the area is covered by a single antenna or more than one (Vandenbulcke et al., 2017). Here we have considered the same error for all HFR radial observations, so as to also exploit the potential benefit of observations in areas covered by only one antenna, as discussed in Shulman and Paduan (2009); Stanev et al. (2015). However, the evaluation of the effect of penalizing observations with larger errors in areas covered by a single antenna or affected by Geometry dilution of precision (GDOP) effects is relevant and would need further research. The observation error should also ideally include correlated observation errors, even if our knowledge of these errors is still somehow limited. This is another interesting aspect that should be evaluated in future studies.

3.5 Conclusions

In this work, we have integrated different multivariate ocean observations with numerical modelling to improve the dynamical knowledge of ocean currents in line with the actual concerns in operational oceanography (De Mey-Frémaux et al., 2019; Kourafalou et al., 2015a,b; Schiller et al., 2015). This work has benefitted from the collaborative framework of the JERICO-NEXT European research infrastructure initiative (Farcy et al., 2019b), which aims at fostering cooperation to build a sustained coastal observing network. We combined high-resolution modelling with satellite and in-situ observing sources and HFR surface currents measurements to discuss the contribution that the developing HFR networks could provide to regional and coastal operational modelling. The impact of HFR-DA has been evaluated, using both radial and total observations along with generic data sources as SLA, SST and Argo TS profiles. The system showed its ability to improve the representation of ocean fields by assimilating different types of observations, from a variety of sources observing a wide range of spatio-temporal scales.

The assimilation of generic observation sources helps to correct surface currents in the IC as revealed by both the Eulerian and Lagrangian validations. The employment of HFR observations further improves the forecasting of surface currents in the IC. While GNR simulations reduce the RMSD and the mean error, the assimilation of HFR leads to an increase in the correlation between model and observations for both components of the velocity. The Lagrangian validation reveals the capacity of HFR data assimilation to improve the prediction of surface currents inside the area covered by both antennas, while not degrading the representation of the more steady currents found outside of it. The experiment assimilating HFR total velocities is the one that best fits the observations. Besides, it provides the best average skill score for Lagrangian prediction and the lowest mean separation distance between drifters and virtual particles. The use of radial observations benefits from the use of an intermediate nudging initialization approach after the analysis. The results presented in this study confirm the usefulness of HFR systems to improve regional operational ocean forecasting models, even when providing a limited coverage with respect to the model domain extension.

Chapter 4

HF Radar Observing System Simulation Experiment in the Ibiza Channel

Abstract

The impact of the expansion of a high-frequency radar (HFR) system in the Ibiza Channel (Western Mediterranean Sea) is evaluated through an Observing System Simulation Experiment (OSSE). The installation of two new antennas in the Iberian Peninsula would complement the existing ones in the islands of Ibiza and Formentera, providing surface currents observations of the full channel. Two different configuration of the same model, validated to give realistic simulations, are used: i) a Nature Run (NR) which we consider as the real ocean state and that is used to generate pseudo-observations, and ii) a Control Run (CR) in which we will assimilate the pseudo-observations. The OSSE is first validated by comparison against a previous Observing System Experiment (OSE). The effect of the new antennas for forecasting surface currents is evaluated in two different periods. The effect of the new antennas is relatively small when the NR and CR depict a similar circulation. However, in situations where both models present higher differences, the error reduction with respect to the use of only the actual system can be of up to 19%. The effects on the transport in the area are also analyzed from a Lagrangian perspective, showing that DA can help to better represent the Lagrangian Coherent Structures present in the NR and constrain the ocean dynamics.

4.1 Introduction

Observations, models and data assimilation (DA) are the three key elements of operational oceanography. Combining them in an optimal way and bridging synergies between the different research communities is key to advance our knowledge of the oceans and be able to answer to societal needs for a sustainable development (Ryabinin et al., 2019; Visbeck, 2018).

In this sense, Ocean Observing Systems (OOS) play a key role, and numerous efforts have been made all over the world to enhance its development and strengthen the collaboration within the scientific community (Sloyan et al., 2019; Moltmann et al., 2019; deYoung et al., 2019). In particular, in Europe, several initiatives have been made or are ongoing to provide better answers to science and to societal challenges (e.g., CMEMS programme, Jerico and Eurosea projects) (Farcy et al., 2019a; Le Traon et al.,

2019; Tintoré et al., 2019).

The rising capabilities of remote sensing and the development during the last decades of *in-situ* observing programs such as Argo (Le Traon, 2013), allowed a better understanding of ocean processes at multiple scales. In coastal areas, Regional Ocean Observing Systems (ROOS) are nowadays providing near real time observations around the globe and they combine observations from moored instruments, periodic cruises, autonomous vehicles, Lagrangian platforms, as drifters, and high-frequency radars (HFR).

Numerical models provide a complete view of the three dimensional structure of the ocean, however, they are inevitably affected by errors from parametrization of non resolved physical processes, discretization issues, or the lack of accurate forcing. To improve reliability, numerical models for operational purposes should be fed with observations through data assimilation. Observing System Experiments (OSEs) assimilating data are performed to evaluate the capability of specific observing systems to correct model forecasts on simulations. Similarly, the potential impact of observing system has to be evaluated to help design these systems. Observing System Simulation Experiments (OSSEs) can be performed to help to optimally design an OOS or a future campaign (Kourafalou et al., 2015a).

OSSEs were first developed for the atmospheric science community, and over time, specific design criteria have been developed to ensure the realism of the assessments performed, as defined in Atlas (1997). In ocean studies multiple OSSEs had been done, however, most of them did not use a full-fledged DA system approach for the evaluation. Generally, Kalman filters, empirical orthogonal functions (EOF) based or different interpolation methods were used to map the observations and reconstruct the ocean state (Guinehut et al., 2004; Ballabrera-Poy et al., 2007; Sakov and Oke, 2008). Following the procedure established for atmospheric studies (Hoffman and Atlas, 2016), Halliwell et al. (2014) applied them for the first time to the ocean, and in the last years, several studies have been done following the criteria exposed in that work, as we will do here. For instance, Gasparin et al. (2019) performed an evaluation of the influence of the future deep Argo float network, and Benkiran et al. (2021) assessed the impact of the assimilation of data from the future SWOT satellite mission in a global-high-resolution model.

In the fraternal twin OSSE approach employed, two models are required: (i) one, hereinafter referred to as Nature Run (NR), which is considered to represent the true ocean and that will be used for validation and to extract the pseudo-observations, and (ii) the model we would like to correct with the assimilation of such pseudo-observations. To be credible, the OSSE should satisfy the following design criteria and rigorous evaluation steps: (a) The models should be validated to give realistic simulations and the pseudo-observations generated in a way that resemble the real ones, including the observation errors, that need to be specifically added. (b) The validation should be performed by comparison to a previous OSE where real observations are assimilated. The same observations should be assimilated in both experiments, except for the fact that, in the OSSE, the pseudo-observations are synthetically generated from the NR. (c) If the impact assessment is consistently the same, we consider the OSSE to be validated.

In the OSSE framework the ocean state is fully known. This permits to assess the impact in regions that normally are not sampled or to experiment additional validation techniques. For instance, we can use Lagrangian techniques for the assessment of transport and the ocean dynamics, such as the Lagrangian Coherent Structures (LCS) obtained from the Finite Size Lyapunov Exponents (FSLE) (d'Ovidio et al.,

2004; Hernández-Carrasco et al., 2011). Ridges of FSLE field reveal LCS, which act as transport barriers. These LCS have been proven to be useful to understand ecological processes, such as nutrients distribution, or oil-spill and search and rescue operations (Hernández-Carrasco et al., 2018; Lekien et al., 2005; Shadden et al., 2005).

Normally, the validation of these Lagrangian techniques is limited. LCS computed from model simulations can be compared to those calculated from geostrophic currents derived from altimetry products (Hernández-Carrasco et al., 2014) which suffer limitations when approaching the coast (Vignudelli et al., 2019; Pascual et al., 2013). Also from HFR measured currents (Hernández-Carrasco et al., 2018), that are limited to cover small coastal areas. The validation of LCS can be performed by comparison with active tracers, as chlorophyll filaments (Lehahn et al., 2007; Hernández-Carrasco et al., 2018), which by their nature can not depict the full structures present in the ocean; SST fronts (d'Ovidio et al., 2004), which are inferred from satellite products that can be affected by clouds; or fish stock concentrations (Baudena et al., 2021), tracked seabirds or marine predators (Kai et al., 2009), that are difficult to monitor. Here we will take profit of the full ocean state knowledge supposed in the OSSE approach. The use of a NR model for the validation of the LCS computed from the model simulations implies a step forward to address the question of how data assimilation can help models to correct the circulation, especially in coastal areas.

This study is a continuation of the work presented in the previous Chapter 4, where a series of OSEs were performed to evaluate the impact of HFR DA on the correction of surface currents in the Ibiza Channel (IC) (Hernandez-Lasheras et al., 2021). In that work, it was shown that using HFR observations together with satellite observations (altimetry and sea surface temperature) and Argo temperature and salinity profiles increases the model's capability to forecast surface currents. In this work we will use the same set-up, using the nudging initialization method described in the mentioned work, as it is the configuration employed in our operational system.

The objective of this study is to evaluate the impact of setting up a couple of antennas in the other part of the IC, in the Spanish mainland, to extend the coverage of the HFR system. We analyze the impact of this new observing system on the transport properties through the LCS computation. The effect of data assimilation on the reconstruction of the LCS and its impact on the spreading of the advected particles is studied.

4.2 Data and Methods

4.2.1 Study Area and HFR system

Our region of study is the Ibiza Channel (IC), in the Western Mediterranean Sea. The modelling area spans from Gibraltar strait in the west to Sardinia and Corsica in the east. The IC is a passage of water between the Iberian peninsula and the island of Ibiza, in the western Mediterranean Sea (Pinot et al., 1994, 1995). It is a choke point between the saltier waters from the north, that generally flow along the coast, and the incoming fresher waters from the south (Heslop et al., 2012).

SOCIB operates since 2012 two CODAR HFR antennas in the islands of Ibiza and Formentera, measuring surface currents in an area up to 80 km far off the coast (Tintoré et al., 2012). In this work we will evaluate the potential impact that a couple of new antennas set in the Iberian peninsula, in the eastern part of Cape la Nao (Fig 4.1) could

have. This way the IC HFR system will expand its coverage area mapping the entire channel. The area, shown in Figure 4.1, is considered as the most likely coverage in terms of total velocities (u-v) that a couple of antennas in the peninsula will provide together with the actual system.

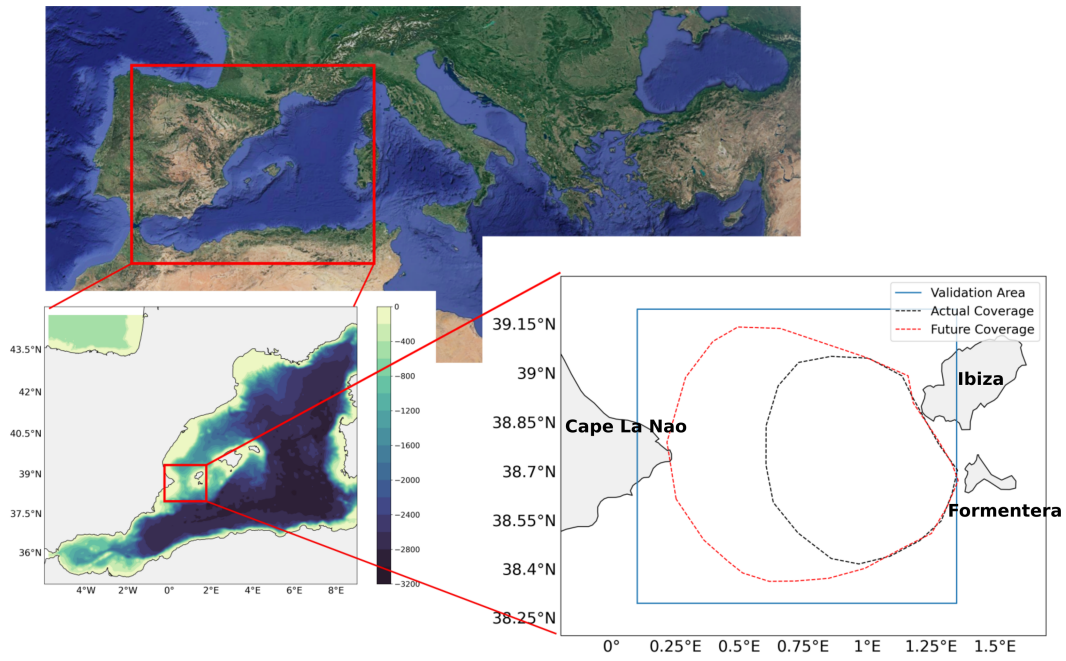

Figure 4.1 *WMOP domain and bathymetry in the Western Mediterranean. HFR actual coverage area (black dashed) and Future one (red dashed) from which observations have been simulated. The area selected for validation is also shown (blue).*

4.2.2 OSSE set-up: Simulations

The OSSE perspective requires two model simulations. Here we used a fraternal twin OSSE system approach (Halliwell et al., 2014), in which two different configurations of the same model are employed: i) a Nature Run (NR), which is considered to represent the true ocean and that will be used for validation and to extract the pseudo-observations, and (ii) a Control Run (CR) model in which we will evaluate the impact of assimilating different observing sources. We use the WMOP (Juza et al., 2016) model, which is a regional configuration of ROMS (Shchepetkin and McWilliams, 2005) for the Western Mediterranean Sea. It spans from Gibraltar Strait in the west to Corsica and Sardinia straits in the east, with a horizontal resolution around 2 km and 32 vertical terrain-following sigma levels (resulting in a vertical resolution between 1 and 2m at the surface). Both the NR and the CR model, which will be used to assimilate, have the same configuration, parametrization and atmospheric forcing, with the only difference in the initial state and the boundary conditions used in each of them. The CR is a free-run hindcast simulation developed and evaluated in Mourre et al. (2018); Aguiar et al. (2020). We will use this model configuration to assimilate data into, as it is the same one used in the reference OSE we will use for validation (Hernandez-Lasheras et al., 2021). It uses the Copernicus Forecasting System for the Mediterranean Sea (CMEMS MED-MFC), with a 1/16º horizontal resolution (Simon-

celli S., 2017), as initial and lateral boundary conditions. NR, in contrast, uses the Mercator Glorys reanalysis global product, with a 1/12 horizontal resolution (CMEMS GLOBAL_REANALYSIS_PHY_001_030) and has also been validated to give realistic simulations, comparing against observational data from satellites and Argo buoys. The atmospheric forcing, common for both simulations, is provided every 3 hours at $1/20^{\circ}$ resolution by the Spanish Meteorological Agency (AEMET) through the HIRLAM model (Undén et al., 2002) and the bathymetry is derived from a 1' database (Smith and Sandwell, 1997).

Both model realizations resolve the same scales, while differing in the mesoscale structures present during the experiment period, which are the two main initial requirements needed for a fraternal twin OSSE approximation. Figure 4.2 shows the Hovmoller diagram of the meridional velocity in a transect across the Ibiza Channel (latitude 38.77N), where we can observe differences between both runs in the currents across the IC during the whole year 2014. The mean circulation pattern in the Ibiza Channel between 21 September and 20 October can be observed in the top two panels of Figure 4.3, where we have also marked (dotted line) the coverage areas of the actual and the possible future antennas considered in this study. Both simulations present in average a southward current in the western side of the IC, while having a northward flow in the eastern part. In the case of the NR both flows are more intense than in the CR, which depicts a more intense eastward current in the southern part of the coverage area.

To further explore the capabilities of the new antenna under different possible circumstances, we selected another period in which the dynamics in the area between both simulations present more differences. For this, we chose August 2014 to repeat our simulations. In the CR simulation, the dynamics during August are similar to the following September-October period, with northward currents in the east side of the channel and southward in the west, as can be seen in Figure 4.2. On the contrary, the NR depicts a northward current in the eastern side of the channel and also in the west, where it is more intense. In the middle of the channel (0.62E-0.85E), there is a strip of weak northward current neither present in the CR. Top two panels of Figure 4.4 shows the mean circulation in the region for the NR and CR.

For the two commented periods we run three data assimilative simulations, using different datasets, and we evaluate the impact comparing against the free-run CR (Table 4.1). We called GNR the simulation assimilating the generic data set, composed of SLA along-track, SST and Argo T-S profiles. H-A and H-F are the simulations that, additionally to this generic data-set, employ HFR simulated total observations from the actual or the future coverage area, respectively.

The data assimilation system employed is the Local Multimodel EnOI scheme previously described that was validated to correctly assimilate HFR observations in Chapter 4. We here use the nudging initialization method after analysis, as it is the one employed in the WMOP operational system and it is less prone to produce discontinuities in the field which could affect the computation of FSLE, that will be later presented.

4.2.3 OSSE Set-up: Pseudo-Observations

For our experiment, the satellite and Argo pseudo observations have been extracted at the same position and time as the real observations in the previous chapter's OSE (Hernandez-Lasheras et al., 2021), that is considered as reference. We simulated along-

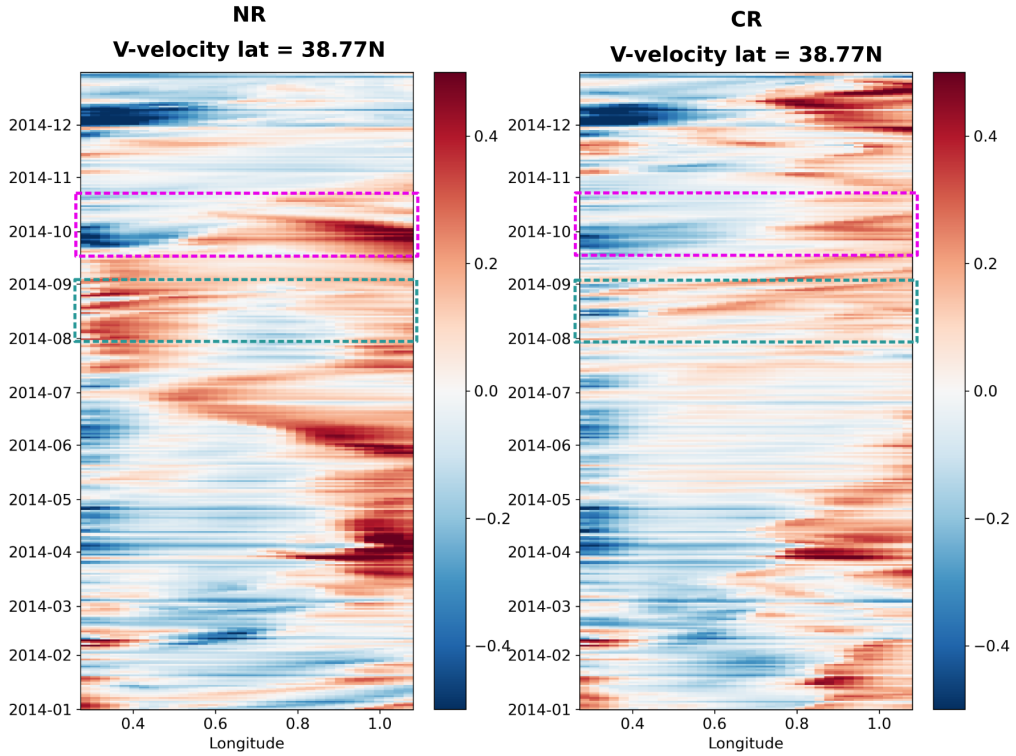

Figure 4.2 *Hovmoller of the meridional component of velocity during 2014 for a transect at 38.77N latitude, in the Ibiza Channel, for NR (left) and CR (right). The two periods of the OSSE are highlighted. The period coincident with the OSE is marked in pink (21 September to 20 October). In green, the second period simulated, during August.*

Simulation	Assimilated observations
NR	**None. Pseudo-reality**
CR	None
GNR	SLA, SST, TS
H-A	SLA, SST, TS, HFR (actual coverage)
H-F	SLA, SST, TS, HFR (future coverage)

Table 4.1 *Basic description of the experiments, indicating the dataset used in the simulations.*

track sea level anomalies (SLA) from four different altimeters (Cryosat, Jason-2, Saral Altika, and HY-2). NR fields are interpolated in space and time to each satellite observation after removing the mean dynamic topography. For SST, we emulate the SST Foundation product, which does not account for the diurnal cycles, sub-sampling surface temperature fields from the NR at 8 a.m., with a 10 km resolution. The Argo profiles were sampled by interpolating the temperature and salinity fields in space and time. We added noise to every observation, which was randomly generated for each observation considering a Gaussian probability distribution with a standard deviation of the value of the error. The observation error has been considered the same as for the real observations. Table 4.2 indicates the value of the representation and instru-

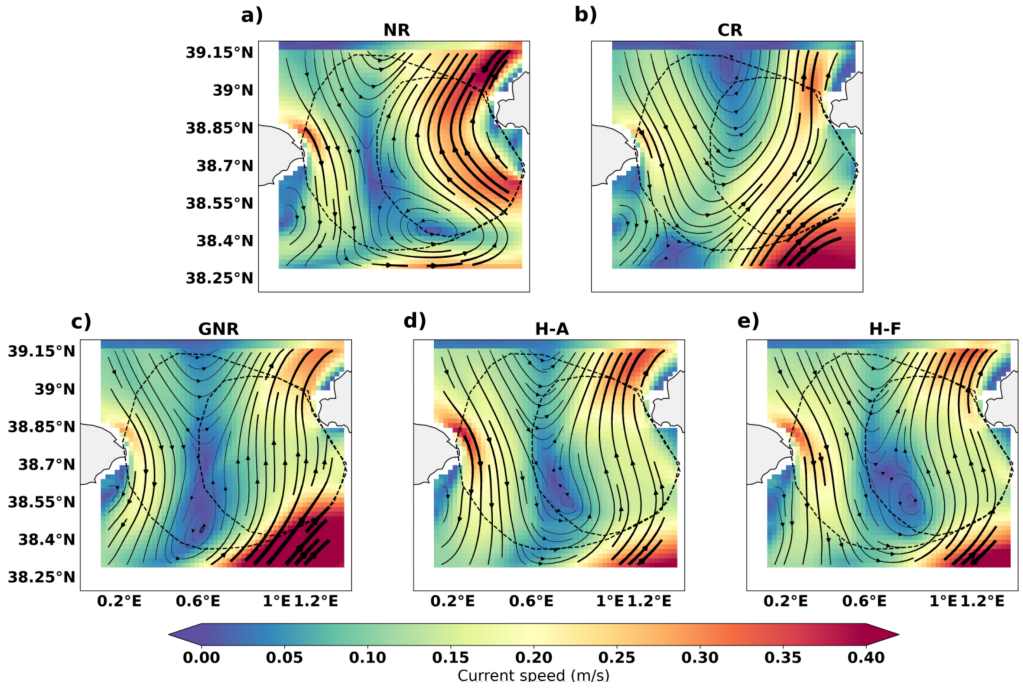

Figure 4.3 *Mean surface circulation during the one-month simulation (20-September to 20-October) in the Ibiza Channel for the a)Nature Run (NR), b) Control Run (CR), c) GNR, d) H-A, e) H-F. Actual and future coverage areas are marked in dashed black lines.*

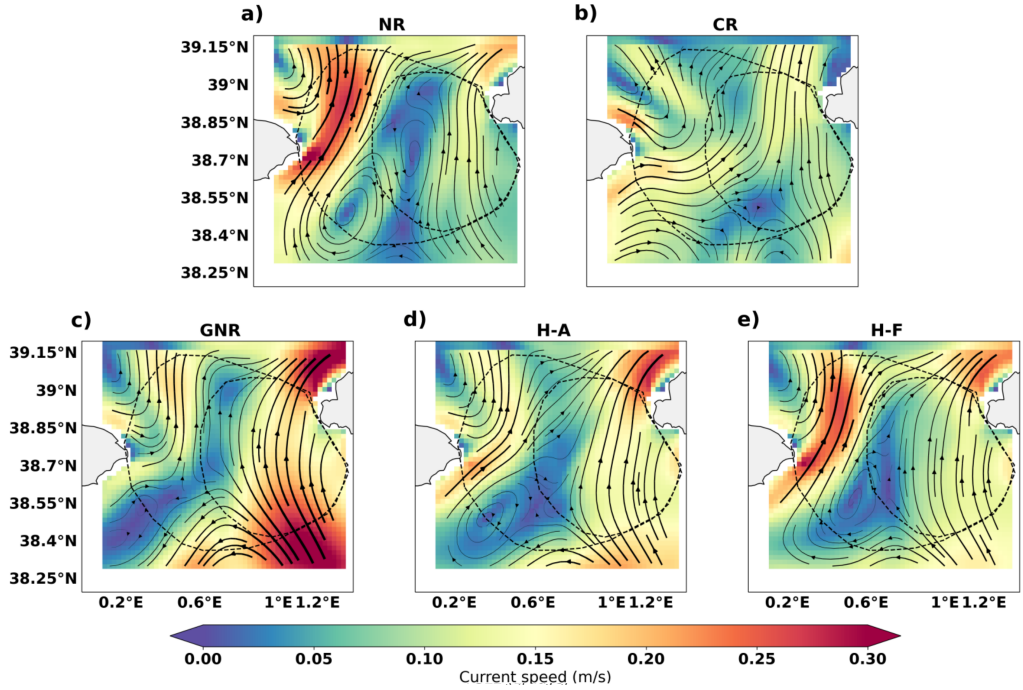

Figure 4.4 *Mean surface circulation during the one-month simulations (August 2014) in the Ibiza Channel for the a) Nature Run (NR), b) Control Run (CR), c) GNR, d) H-A, e) H-F. Actual and future coverage areas are marked in dashed black lines.*

	Representation	Instrumental
SLA (m/s)	0.03	0.02
SST (º)	0.25	0.5
Argo - T (º)	0.25	0.1
Argo -S	0.05	0.01
HFR (m/s)	0.03	0.02

Table 4.2 *Representation and instrumental errors employed for the different observations.*

	SLA (cm)		SST (ºC)		Argo T (ºC)		Argo S	
	mean	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**
OSE	0.07	0.05	0.01	0.62	-0.15	0.90	-0.00	0.21
OSSE	0.01	0.06	-0.01	0.78	-0.02	0.81	-0.03	0.19

Table 4.3 *Mean value and standard deviation of the Innovations of SLA, SST and Argo T-S for the OSE and OSSE.*

mental errors considered for the different observations. For Argo observations only the horizontal representation error is shown. Note that the total error has the expression $\sigma^2_{tot} = \sigma^2_{rep} + \sigma^2_{ins}$.

Figure 4.5 shows the histogram of the innovations (observation - model) for the OSE (blue) and OSSE (red), whose results are also synthesized in Table 4.3. We can observe that all observations follow a similar distribution both for OSE and OSSE. The only discrepancy is found in the SLA, where we can observe a bias of 0.07 m in the OSE. This is a known issue concerning the value of the satellite SLA observations. Satellite SLA observations are computed by extracting the value of the absolute dynamic topography measured by the satellite to the mean dynamic topography (MDT). The MDT of the model generally differs from the MDT of the observations. Moreover, two different simulations have two different MDTs. The rise of the mean sea level generates a climatological trend in the observations which has not been corrected and thus, SLA observations from recent years present a positive bias which also affects the innovations. However, we believe this does not significantly affect the assimilation of SLA, as it is not corrected during the simulation (as discussed in the previous chapter (Hernandez-Lasheras et al., 2021)) and do not impact the geostrophic circulation. The values of the innovations standard deviation, which is directly related to the centered-root-mean-square-deviation (CRMSD), has small differences between OSE and OSSE for all observing sources.

For the HFR observations we have followed a slightly different approach. We have considered two polygons, one containing the actual coverage area and another considering the potential future coverage that a set of two antennas settled in the western part of the Ibiza Channel might provide, according to expert criteria (Fig. 4.1). Within these areas, we have sub-sampled the daily mean velocity fields of the NR in a spatial grid of 3 km resolution, which corresponds to the HFR total (u-v) observing resolution in the area (Tintoré et al., 2012). We randomly dismissed 15% of the observations for simulating potential gaps in the antennas coverage. Again, we have introduced a Gaussian noise to the observations and considered the same instrumental and representativity error we previously used with the real data. Figure 4.6 shows the histogram of the innovations

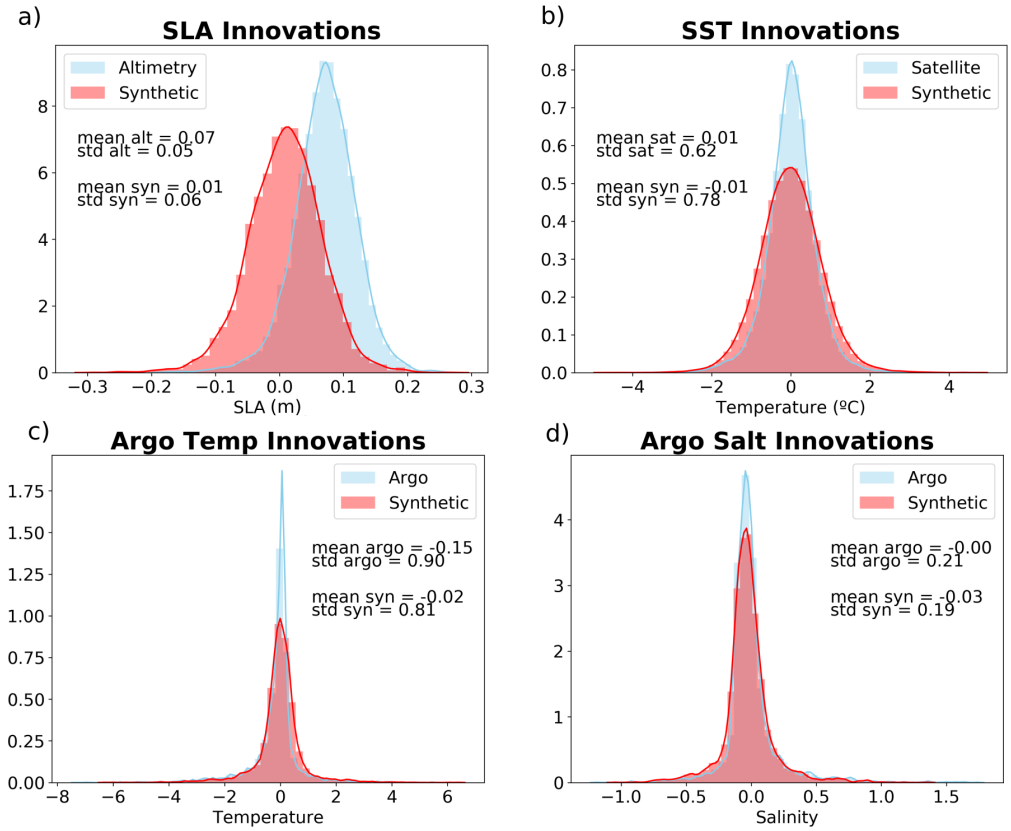

Figure 4.5 *Histograms comparing the innovations (observation - model) for OSE (blue) and OSSE (red) for the different observing sources: a) SLA along track, b) SST, c) Argo temperature profiles, d) Argo salinity profiles.*

for both variables of velocity during the two periods run.

It can be observed that during the period of 21 September to 20 October, coincident with the previous OSE, the innovations present a larger discrepancy, both in mean and standard deviation, as seen in the spread of the distribution. In this period, the model tends to overestimate the meridional currents observed by the HFR. However these discrepancies are not systematic, as can be seen for the period of August, where the innovation distribution is much more similar when using real or virtual observations. While the meridional component still has a mean difference, the standard deviation is almost equal whether using real or virtual observations for both velocity components.

Overall, the statistical properties of the innovations are consistent between the OSE and the OSSE, which validates the use of the NR to generate pseudo-observations. The validation of the OSSE framework employed will be further completed, assessing the impact of the observations on the model.

4.2.4 Lagrangian Analysis. Finite size Lyapunov exponents (FSLE)

The effects of the field in the transport are addressed by a Lagrangian description. The Lagrangian perspective has the advantage of exploiting both spatial and temporal variability of a given velocity field (Hernández-Carrasco et al., 2011). In the finite size Lyapunov exponents (FSLE) technique a set of fluid particles, initially separated a distance δ_0, are followed in time as they are transported by the flow, integrating the equation of motion. The FSLE, denoted by λ (Eq. 4.1), is inversely proportional to the

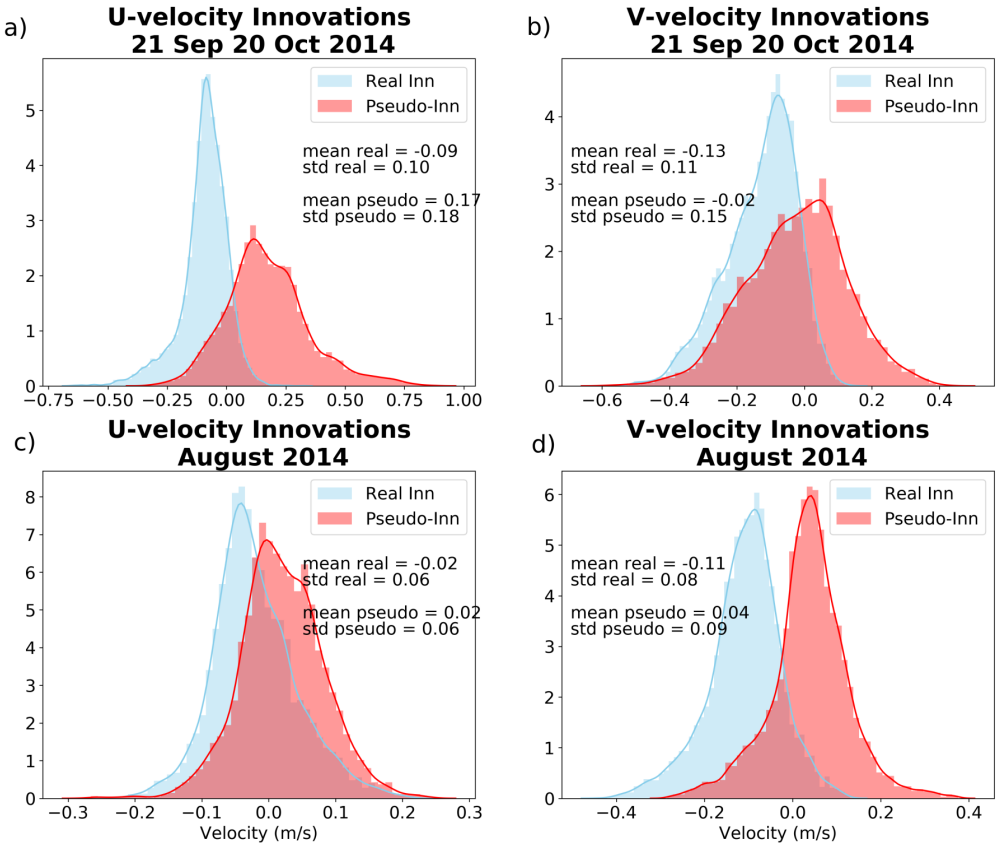

Figure 4.6 *Histograms comparing HFR Pseudo-obs created from NR against real observations assimilated in previous OSE: a) zonal velocity, b) meridional velocity.*

time at which two fluid particles initially separated δ_0 get separated a certain distance δ_f. When integrated backwards, high values of FSLE are identified as limits of maximum stretching. Lagrangian Coherent Structures (LCS) are defined as ridges of FSLE fields, which act as transport barriers.

The integration of the trajectories is performed with OceanParcels (Lange and Van Sebille, 2017), using a Runge-Kutta 4 algorithm with an integration time step of 1 hour. The initial separation distance δ_0 is considered equal to the model grid resolution (around 2 km). The final distance $\delta_f = 10 \cdot \delta_0$, as considered optimal in other studies which explore the importance of this parameter (Hernández-Carrasco et al., 2011; Hernandez-Carrasco et al., 2012).

$$\lambda\left(\mathbf{r}, t, \delta_0, \delta_\mathrm{f}\right) \equiv \frac{1}{\tau} \log \frac{\delta_\mathrm{f}}{\delta_0} \tag{4.1}$$

We calculated the FSLE field in the entire domain, launching particles with an initial separation equal to the model grid size. For the one-month simulations, we computed the FSLE integrating the trajectories backwards in time during 15 days. From the 16th simulation day onward, the last 15 days' fields are used to obtain one FSLE field.

4.3 Results

4.3.1 OSSE validation: Comparison with OSE.

First, we assess the results of our experiments comparing them to the ones obtained in the previous chapter's OSEs used as reference (Hernandez-Lasheras et al., 2021). When the forecast error reduction obtained with both perspectives (OSE and OSSE) is similar, we can consider the system is valid, and reliably estimates the impact of other potential observations. This would enhance our reliability in the assessment of the future observation system design. For this validation we analyze the period spanning from 21 September to 20 October from an Eulerian perspective, in a similar way as it was done in the previous OSE: a) For SLA, we compare for each day of simulation the model equivalents against the NR at every location of all along-track possible observations in the region; b) SST comparison between model and NR is performed at every observation point within a grid of 10 km resolution, like the one used to generate the pseudo-observations; c) Temperature and salinity fields are interpolated in time and space to the Argo float profile observations; d) For comparison against HFR data we interpolate the surface average fields to the position of the real observations. Note that in the OSSE perspective, we compare against the value of the true ocean state, represented by the NR simulation. Thus, although the evaluation with the NR is done at some of the same points from where we extract the assimilated observations, the validation can not be considered as fully independent.

Table 4.4 shows for each observing source the CRMSD normalized with the CR for the OSE and OSSE. This metric give us an overview of the impact, without taking into account the mean error (bias), which is only significant in the case of SLA observations, as was commented previously (section 4.2.3). For SLA, SST and Argo T-S only the results of the GNR simulation are shown, as the ones obtained for H-A are almost the same when comparing against these sources. For the comparison against HFR data we show the results of GNR and H-A simulations (Note that H-A corresponds to TOT-N simulation in the Chapter 3). We can observe that the normalized CRMSD for GNR is slightly higher for the OSSE than for the OSE, with even an increase in the error for the v-component. On the contrary, H-A produces better results for the OSSE in the u-component. This suggests a bigger impact of adding HFR observations in the OSSE approach for this specific period. For the rest of observing sources we can observe that the error reduction between OSE and OSSE is of the same order.

A further analysis of the results is shown in the Taylor diagrams in Fig. 4.7. For SLA, the model error decreases around 40% and the correlation between model and observation increases from 0.38 to 0.75 when using DA. Results are almost equal for all three simulations using DA. For SST, the comparison between model and NR is performed at every observation point within a grid of 10 km resolution, like the one used to generate the pseudo-observations. Results show how the error reduction is around 36%, while correlation increases for all simulations from 0.87 to 0.94. These results are of the order of magnitude of the ones obtained for real observations (shown in chapter 3). Similarly, for the Argo T-S observations, the results obtained for the OSSE are very similar and of same the order of those obtained with real-world observations. The error is reduced by 41 and a 36 %, for temperature and salinity respectively, and the correlation increases in both cases (Fig. 4.8).

		SST	SLA	Argo-T	Argo-S	HFR-U	HFR-V
OSE	GNR	0.69	0.73	0.66	0.62	0.80	0.96
	H-A					0.71	0.74
OSSE	GNR	0.64	0.60	0.59	0.64	0.87	1.06
	H-A					0.64	0.74

Table 4.4 *Comparative table between OSE and OSSE experiments, of the normalized RMSD against the different observing sources. SLA, SST and Argo values are only shown for GNR simulation in each case, as the values for the H-A are the same ones.*

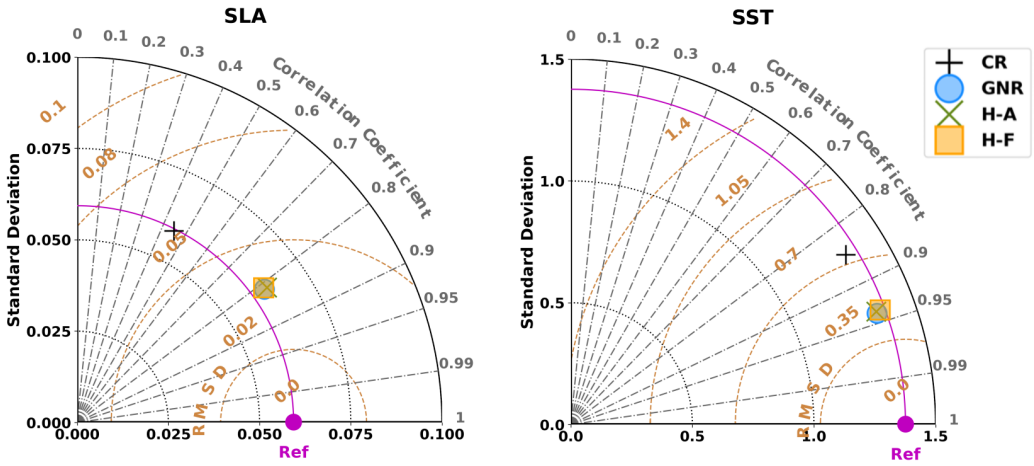

Figure 4.7 *Taylor diagrams comparing models and NR pseudo-observations in terms of SLA (left) and SST (right) over the whole modelling domain. X and Y axis represent the standard deviations of the data. Distance from the reference point located on the X axis (noted as Ref. in magenta) represents the centered root mean square deviation (CRMSD). Correlation between observations and model increases clockwise. Symbols represent the different simulations, as specified in the legend*

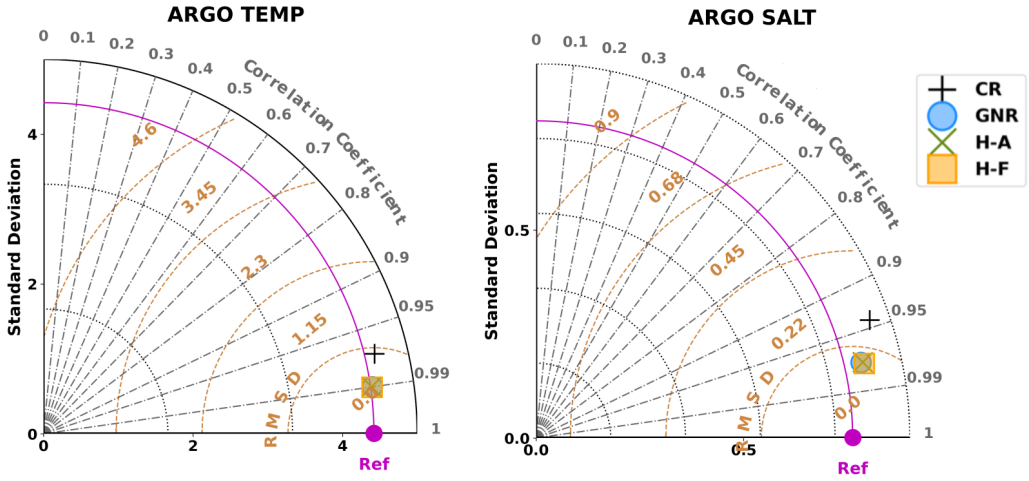

Figure 4.8 *Same as Fig. 4.7 for Argo salinity and temperature*

4.3.2 Impact of the HFR system expansion. Eulerian validation

We further assess the impact of OSSEs on the surface currents comparing the model fields in each simulation against the NR in the area of the IC. Note that the approach is

different to what we previously did to validate the OSSE by comparing it to the OSE as only the data at the observing points during the coincident period were evaluated previously. The area used for the assessment in this section, which can be seen in Fig. 4.1, covers the entire IC, being wider than the coverage of the future HFR system. This way, we evaluate the impact of HFR DA beyond the coverage of the antennas and its effects on the transport in the region.

We evaluate the two different periods of simulation. During September-October, the meridional currents in the region are more intense, as can be seen in the Hovmoller diagram 4.2 and so are the errors. The assimilation of generic observations only does not improve the prediction of surface currents in the region. For the u component, GNR improves the correlation with the NR from 0.15 to 0.43 but only slightly reduces the CRMSD (centered root mean square deviation). For v, both the correlation and error are slightly degraded in comparison to CR. The use of HFR observations additionally to the generic sources is here essential to improve the forecasting of both zonal and meridional components. The improvement obtained when assimilating observations from the future HFR system is slightly better than that obtained when only using observations from the actual coverage area. Both for H-A and H-F the correlation are higher than 0.62, and the error is reduced by more than 32% for the u-component. For the v-component, the correlation for both simulations also increases with respect to the NR, and the error is reduced by 15% and 21% for H-A and H-F, respectively.

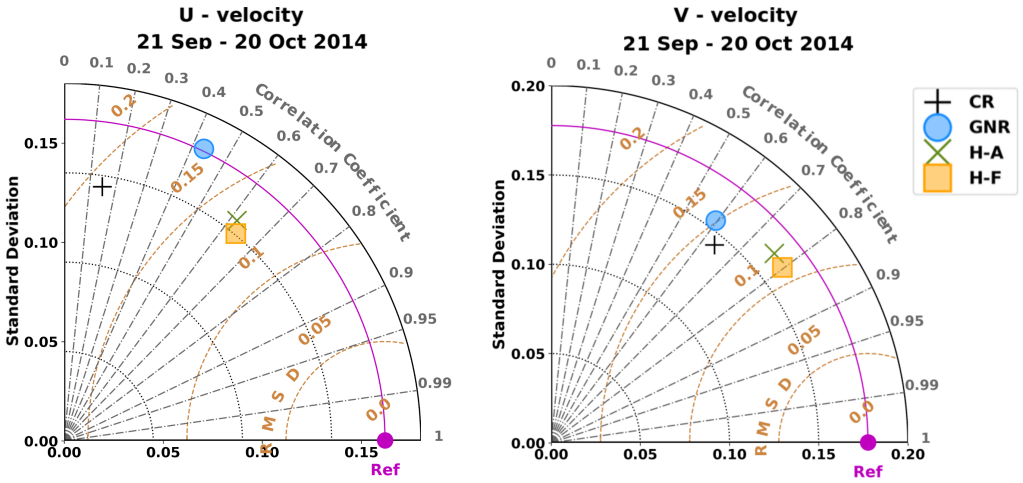

Figure 4.9 *Same as Fig. 4.7 for velocities in the IC for the simulations spanning 21 September and 20 October 2014.*

During the second period (1-30 August), the meridional currents were less intense than for the other simulation period, both in NR and CR. Furthermore, in NR, the currents in the western part of the IC were northward, contrary to CR and the area's mean dynamics. This in an anomalous situation circulation in the area, but that is known to happen few times every year. We aim to explore the potential impact of the HFR system extension in such situations, where the model could differ more from the observations. GNR degrades the forecasting of surface currents compared to CR. HFR observations are also needed to reduce the error and increase the correlation between the model simulations and the NR. When using HFR observations for the zonal velocity the correlation increases from 0.50 for the CR to 0.76 and 0.80, and the error is reduced by 23 and 29%, for H-A and H-F respectively.

The difference between using observation only from the actual coverage area and from the future one is significant for the meridional velocity. While H-A increases the correlation from 0.39 to 0.57 and reduces the RMSD by 17%, H-F further reduces the error by 19% with respect to H-A, meaning a total 32% total error reduction, with a correlation of 0.71.

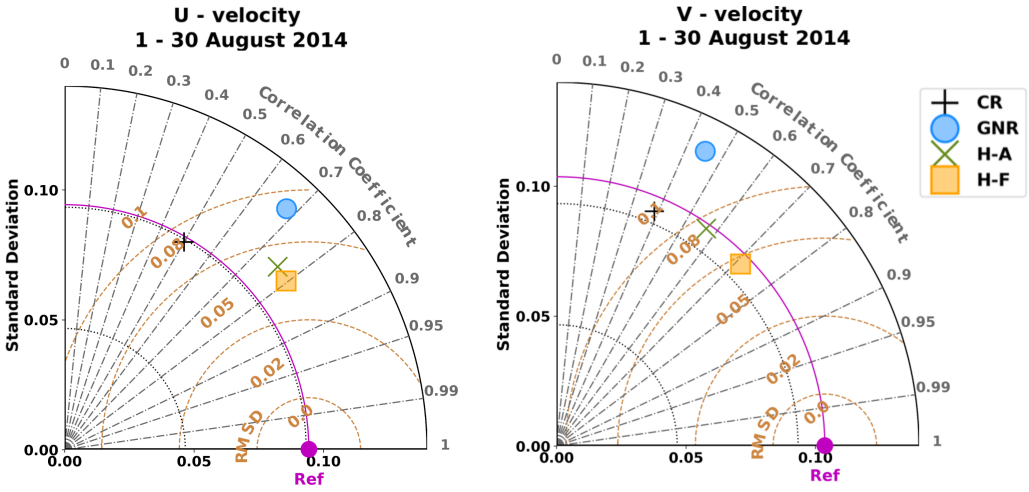

Figure 4.10 *Same as Fig. 4.7 for velocities in the IC for the simulations spanning 1 and 30 August 2014.*

4.3.3 Lagrangian validation

We focus our study on the region surrounding the IC, checking if the model can reproduce the LCS present in the NR when using DA. A qualitative analysis reveals that DA changes the LCS of the model with respect to the CR and can generate some of the structures present in the NR.

The LCS show areas of particle accumulation and barriers to transport. To better understand how DA impacts the dynamics in the area and how the transport patterns can be modified in the IC, we launch every 3 hours a set of 1000 particles in 4 different regions: i.e., north, south, east, and west of the IC. The regions are selected based on a geographical situation to evaluate the zonal and meridional flow exchanges.

Figure 4.11 shows the FSLE field for October 14 and the position of all the particles launched at the four sites every three hours since eight days before. The main LCS significantly differ between NR and CR in all the domain. CR shows an eddy in the southwest, centered at 1E 37.6N, that traps particles (in red) deployed at the south, while in NR, we can observe a loop-shape structure southwards. Red particles in NR move northeastwards between two LCS that make all particles flow uniformly until they arrive east of Ibiza island, where the field is less steady and with more diffusion, driving some of the particles southwards. This behavior is reproduced in the simulations using DA. H-F is the one that better reconstructs the LCS obtained with NR velocity fields.

The LCS present in the middle of the Ibiza Channel in NR are also well reproduced in both simulations where HFR data are assimilated. The orange particles flow eastwards until reaching this LCS, which acts as a barrier, splitting the possible track of the particles in two branches surrounding the island of Ibiza. This situation is not reproduced in CR, where all particles flow eastwards towards Ibiza crossing to the north side

after a few days.

For the blue particles deployed in the eastern side of the IC we can observe how in NR most particles spread around Ibiza. For CR, the particles are quickly advected north-eastwards reaching the north of Mallorca island after a few days, following the LCS that joins the east parts of both islands. This structure is not so intense but still present in the data-assimilative simulations. However, most of the particles still remain close to Ibiza.

Finally, the set of particles deployed at the north (brown) are more dispersed in the DA simulations, along the LCS that is formed north of Ibiza at 39.2 N approximately. This structure is present in NR, but the particles slightly move during the eight days of simulation.

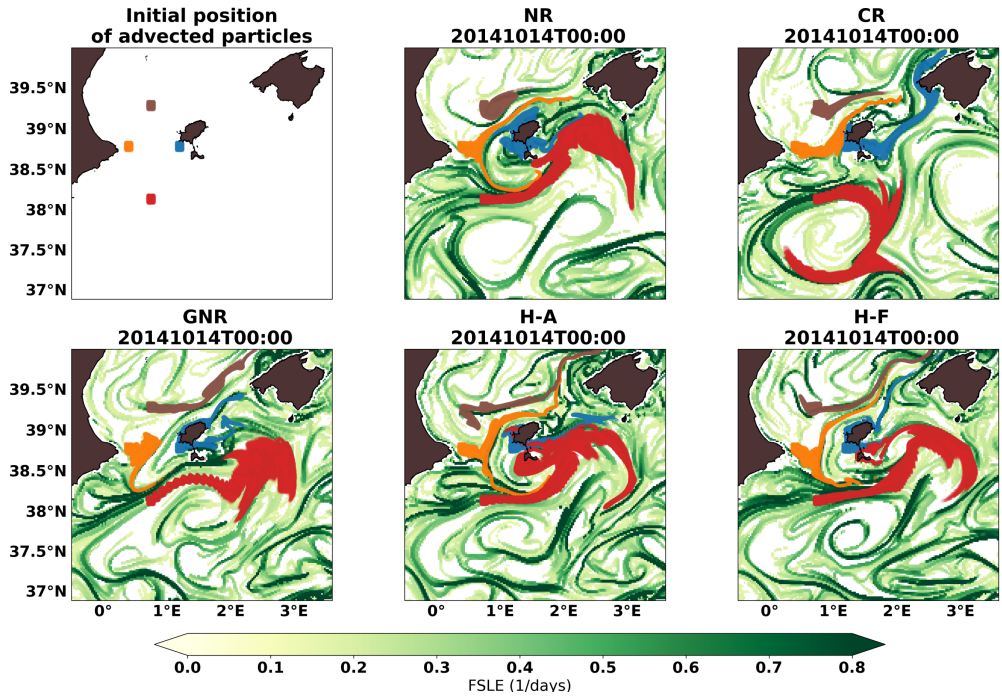

Figure 4.11 *Lagrangian coherent structures from FSLE calculated backwards corresponding to 16-October-2014. Particles launched every 3 hours, starting in 6 October, from 4 different areas surrounding the IC are also shown with different colors.*

Figure 4.12 depicts the FSLE fields for August 24, 2014 and the position of the particles, which were continuously launched every three hours since up to 8 days before. NR presents two big round shaped LCS in the south and east part of the plotted area, probably due to two respective eddies. CR also presents two big structures, but more displaced to the east. The zonal transport in the Mallorca-Ibiza channel will be restricted in the CR by a LCS that extends along the north coast of Mallorca and crossing the channel southward. On the other hand, the motion of particles is constrained meridionally in the IC, especially in the simulations with DA. In NR, the northern part of the eddy previously described would block this transport, while several structures limit it in all the DA simulations.

The most relevant difference found are regarding the transport of particles between the different simulations is seen in the northern and eastern sides of the channel. As seen in NR, orange particles flow northwards, as expected by the mean current seen during

this period (Fig. 4.4), until they get blocked by a LCS, limiting the transport of brown particles at the south, that extend along the LCS in both directions. This behavior, that is not well reproduced in CR, can be reasonably well captured in H-F, where particles motions depict a very similar pattern, and also for H-A, but to a less extent. In this simulation, there is a slight displacement of some orange particles southwards at the initial stages, and we do not identify the left branch of the orange particles flowing northwards, as in NR and H-F.

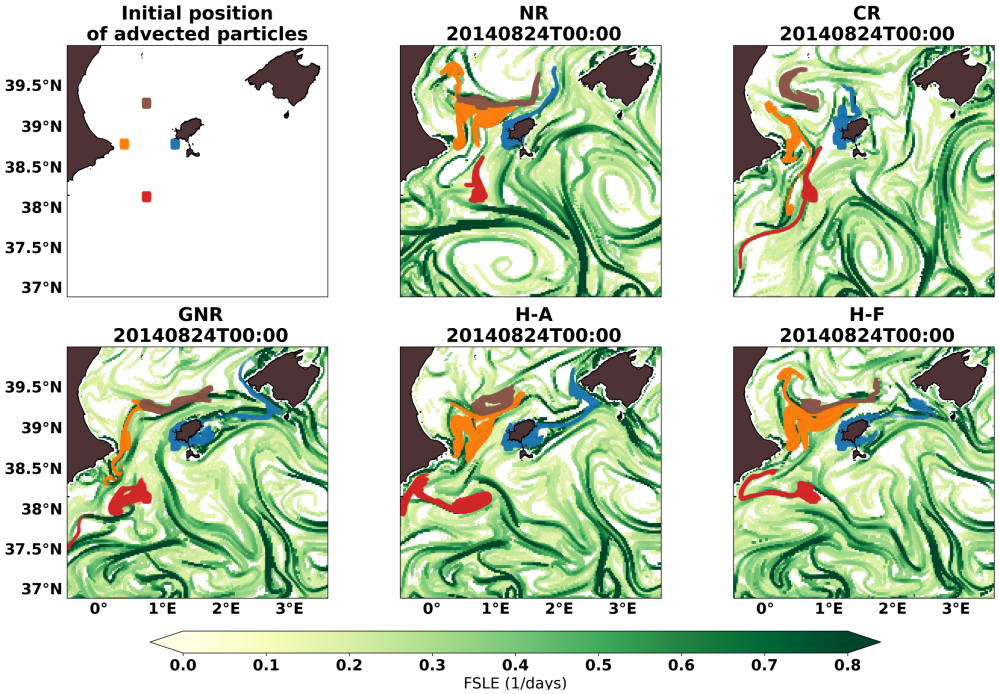

Figure 4.12 *Lagrangian coherent structures from FSLE calculated backwards corresponding to 24-August-2014. Particles launched every 3 hours, starting in 16 August, from 4 different areas surrounding the IC are also shown with different colors.*

4.4 Discussion

The experiments presented here apply an approach which has not been applied before for the design and extension of a HFR system. The NR is validated to give realistic simulations and the innovation distribution of the pseudo-observations present a similar distribution to the real data. The results obtained in the OSSE framework in terms of error reduction and correlation increment are of the same order as of the ones obtained in the previous OSE work. For surface currents, there is a significant difference between the innovations obtained in OSE and OSSE, however we have seen that this difference is not systematic and depends on the analyzed period (Figure 4.6). Furthermore, when assimilating HFR, the error reduction and the increment in the correlation are also of the order of the one achieved in the OSE.

We have evaluated the impact on the surface currents in a wider area than that covered by the antennas, taking advantage of the full ocean state knowledge provided by the OSSE framework. It draws attention that the assimilation of only generic observing sources cannot correct the circulation in the area in the OSSE. In the previous OSE the

assimilation of generic sources led to a good improvement compared to the CR. Even though the model's response to SLA, SST and Argo observations is very similar to that obtained with real observations, the circulation seems to have a highly ageostrophic component. Therefore, the correction of SSH and density fields is not able to correct the surface currents in this region. This enhances the need of high-resolution surface current observations in coastal areas where satellite-derived geostrophic currents tend to fail (Vignudelli et al., 2019).

The extension of the HFR system seems to be useful under certain conditions. The new antennas provide a moderate effect when NR and CR simulations reproduce similar circulation patterns. However, when both simulations present more differences, especially in the western side of the channel, having surface current observations of the full channel is key to improve the model dynamics. This can be interpreted as real situations in which the model is unable to reproduce the dynamics observed by the HFR system.

The observation error for the HFR has been considered the same in all the domain, analogously to what we did in the previous Chapter 3. It could be supposed that in areas covered by three antennas more radial observations will be used to generate a total observation, reducing the expected error. This approach, using a spatial variable error depending on the number on antennas that cover an area has been explored by some authors (Vandenbulcke et al., 2017). The improvement of the observations error, including correlated errors remains as a potential aspect to explore in future studies. Besides, the generation of all pseudo-observations could be made more realistic by generating the random noise based on a spatial structure, using a given correlation length.

We use here a Lagrangian approach to evaluate the impact of data assimilation on the surface transport. The Lagrangian techniques, such as FSLE, have been increasingly being used in the last years. However, the effects of using a model field sequentially corrected through DA remains still poorly studied. As particles are advected, the DA simulation's discontinuities might impact the trajectories of the particles and the following computation of FSLE. Here we showed that simulations using DA behave in a similar way to those without DA. The particles tend to accumulate along LCS, which act as barriers to transport. The possible discontinuities do not seem to affect or create artifacts in the FSLE field or LCS. When comparing the FSLE fields computed for consecutive days, the transition between them is smooth. There is no significant difference when comparing the variation of LCS for two consecutive days with or without DA between them. Furthermore, the experiments performed here show how we can reconstruct some LCS present in the ocean state when assimilating observations. In particular, the use of HFR data for assimilation helps to recreate the LCS present in the NR and to correct the dynamics and the transport in the region, as was demonstrated with the advection of particles.

The four different areas in the IC from which we continuously deployed particles were selected in terms of their geographical location in the IC to evaluate the zonal and meridional transports in the channel and the connectivity between the different regions. The study could also be complemented by analysing the trajectories of particles launched at different sites as the ones shown in this work. Besides, a further quantitative analysis would be desirable, although, establishing a metric for this kind of analysis is difficult and not extended in the literature. FSLE fields should not be compared pointwise, as little differences in the position of LCS could affect the results, leading to an erroneous interpretation. Developing a valid metric to quantify LCS differences remains as a future work.

OSSEs are an important tool to explore the capabilities of a future observing system design. Strictly, in scientific terms, it is always good to have as many observations as possible. However, as resources are limited, synergies between observing and modeling communities are needed to benefit mutually, and observing systems should rigorously be designed to meet user requirements and respond to societal needs (Davidson et al., 2019). For an optimal design of the observing system expansion further considerations could be taken into account. For instance, different locations for the antennas may be examined. For a final design, the decision should be jointly based on the scientific, and on the logistic and economic assessment, that are out of the scope of this work.

4.5 Conclusions

A series of OSSEs assimilating HFR data along with traditional observing sources (SLA, SST, Argo) is presented here. The study is a continuation of the work presented in the previous Chapter 4, where an OSE (Hernandez-Lasheras et al., 2021) that is here used as reference to validate the OSSE framework was performed. The assessment of the OSSE is consistent with that of the reference OSE and the framework is considered suitable for the design and evaluation a future observing system.

The impact of a potential extension of the actual HFR system in the IC has been assessed. The two new antennas would provide a full coverage of the surface currents in the IC and could help to improve the forecasting of the circulation in the region. In circumstances where the flow regime represented by the model disagrees with the observed one, a diminution of up to 19% of the error can be expected when assimilating the future system observations, compared to the actual ones.

A Langrangian analysis based on FSLE revealed that the use of model outputs corrected with DA are useful for this kind of analysis and are not significantly affected by possible field discontinuities. Furthermore, the analysis showed how the assimilation of HFR observations can help to reconstruct the LCS present in the NR and constrain the circulation in the IC.

Chapter 5

Summary, conclusions and future work

This Thesis deals with some of the actual concerns in operational oceanography, and more specifically, with the improvement of ocean modelling forecasting using assimilation of data from relatively recent ocean observing platforms. In operational oceanography, models need to be merged with real-time observations in an optimal way, using data assimilation, to improve the realism of the simulations and properly answer to scientific or societal demands. The main objective was to evaluate how the assimilation of new observing sources, such as gliders or high-frequency radars, impact the predictions of a high-resolution regional model. The study is focused in the Western Mediterranean Sea, where a series of different data assimilation experiments were performed. In the different chapters, a series of Observing System Experiments (OSE) and Observing System Simulation Experiments (OSSE) are presented to evaluate the impact of different data observing sources such as gliders and high-frequency radars (HFR) to correct the WMOP system developed at SOCIB. WMOP is a 2km high-resolution regional model which spans from Gibraltar strait to Corsica and Sardinia in the East. A Local Multimodel Ensemble Optimal Interpolation data assimilation scheme was developed for DA. The system can ingest both dense local observations and larger scale observing sources from the Global Observing System, as satellite altimetry and sea surface temperature, and Argo float temperature and salinity profiles. Different initialization methods to reinitialize the model after analysis have been explored. We have also employed different observing sources for the validations, such as Scanfish or surface drifters, besides several assessment techniques, like field reconstruction using variational methods as DIVAnd or different Lagrangian techniques.

Chapter 2 explores the performance of different sampling strategies based on either a dense CTD survey or a glider fleet for improving model forecasting capabilities in a specific area. The experiments were performed in the context of the REP14-MED experiment, led by CMRE (Centre for Maritime Research and Experimentation) and supported by 20 partners from six different NATO nations. The sea trial was carried out in June 2014 off the west coast of Sardinia and designed to assimilate intensive campaign data from CTDs and gliders, along with satellite SST and SLA, as well as Argo profile observations.

We used the local multi-model EnOI scheme, following 3-day recursive cycles with a 1-day nudging initialization phase after analysis. Six simulations were performed, assimilating observations from the sea trial, together with generic observing sources (SLA, SST and Argo profiles) all along the domain. Simulations include assimilation of one, two, three, four or eight glider platforms, or alternatively, CTD casts collected by two research vessels during the first Leg of the campaign. The eight gliders made

back and forth transects in parallel across the study area, with a separation of about 10 km between them, similar to the meridional spacing of the CTD casts. The impact of the campaign data was assessed comparing to a simulation assimilating generic observing sources only and to a free run without assimilation. Observations from CTD and Scanfish collected during the Leg 3 of the campaign were used for independent validation. The DA system was shown to perform correctly, allowing to properly ingest both large-scale data over the whole western Mediterranean domain and high-density temperature and salinity profiles collected during the sampling experiment over a limited area. These improvements persist during the 3-day periods separating two analyses.

The assimilation of SLA, SST and Argo profiles leads to an average 15% error reduction when compared against these independent observations, which is further reduced when adding dense CTD or glider observations from the campaign. The error progressively reduces as the number of gliders sampling the area increases. The average error reduction is 24% when using observations from a single platform, and 28%, 33%, 35% and 40% when considering two, three, four and eight gliders, respectively. Assimilation of CTDs only during the first Leg of the campaign outperforms the results obtained with the four-glider configuration, leading to very similar results in terms of RMSD to the eight-glider fleet configuration. The 10 km spacing offered by both sampling strategies is essential to improve the representation of the mesoscale variability in the study area. However, some limitations were found associated with the smoothing effect of ensemble covariances, which do not allow to exactly represent the smaller-scale features present in the high-resolution observations. Results evidence the use of gliders as a good alternative to traditional CTD surveys due to its capability to monitor in all weather conditions, work in difficult access areas and the cost reduction it implies. Moreover, the performance from a glider fleet could certainly provide increased performance with the use of adaptive sampling procedures to optimize the regular track.

In Chapter 3, the objective was to evaluate the impact of assimilating surface currents measured by a HFR in the Ibiza Channel. We combined high-resolution modelling with satellite and in-situ observing sources and HFR surface currents measurements to discuss the contribution that the developing HFR networks could provide to regional and coastal operational modelling. The impact of HFR-DA has been evaluated, using both radial and total observations along with generic data sources as SLA, SST and Argo TS profiles. Different initialization methods after the analysis have also been explored. The system, assimilating different types of observations from a variety of sources observing a wide range of spatio-temporal scales, showed its capacity to improve the representation of ocean fields.

The assessment has been done from Eulerian and Lagrangian perspectives, for which we used a set of 13 drifters deployed in the area during the experiment. Both the Eulerian and Lagrangian validations revealed that the assimilation of generic observation sources helps to correct surface currents in the IC, leading to a reduction of the RMSD and the mean error. The forecast capacity of the system is further improved when employing HFR observations, providing an increase in the correlation between model and observations for both components of the velocity. Assimilation of HFR data improves the prediction of surface currents inside the area covered by both antennas, while not degrading the representation of the more steady currents found outside of it, as revealed by the Lagrangian validation performed. The simulation assimilating HFR total observations provides the best results, with the lowest mean separation distance between drifters and virtual particles and the highest average skill score for Lagrangian pre-

diction. While radial observations give a lower performance than totals when restarting directly from the analysis, they benefit from using an intermediate nudging initialization approach after the analysis. The results presented in this study confirm the usefulness of HFR systems to improve regional operational ocean forecasting models, even when providing limited coverage with respect to the model domain extension.

Finally, in Chapter 4, a series of OSSE assimilating HFR data along with traditional observing sources (SLA, SST, Argo) are presented. This study can be understood as a continuation of the one done in the previous chapter. The possible future expansion of the Ibiza Channel HFR system is evaluated. A couple of new antennas, set in the Spanish mainland, would provide full coverage of the surface currents in the channel. A rigorous validation of the OSSE framework is performed, comparing against the previous chapter's OSE used as reference. The assessment of the OSSE is consistent with the reference OSE, and the framework is considered suitable for the design and evaluation of a future observing system.

Under certain dynamical conditions, the expansion of the HFR system coverage could provide a significant improvement with respect to the actual one, with a diminution of the error of up to 32%. Besides, it was demonstrated that the simulations using data assimilation can be used to perform a Lagrangian analyses based on FSLE computation without being affected by possible field discontinuities. Furthermore, the use of HFR observations helps to reconstruct the LCS present in the Nature Run, improving the transport and the ocean dynamics in the Ibiza Channel region.

As a general conclusion, the results of this Thesis answer to the questions initially proposed, helping to develop our modelling capabilities and highlighting the importance of combining high resolution coastal observations as gliders and high-frequency radars with traditional observing sources. Combining all kind of observations improves the forecasting and leads to a better understanding of the ocean dynamics and the processes that occur in coastal areas and in the transition between coastal and open ocean regions. Some of the results obtained through this work have helped to develop the SOCIB forecasting system, and are actually used to provide improved operational response, for instance, for search and rescue operations, to understand ecological processes or to give support to field campaigns. Besides, the outcomes of this Thesis have contributed to several research works: e.g., Onken et al. (2018); Mourre et al. (2018); Aguiar et al. (2020).

The assimilation scheme refinement, improving the representation of errors, and including uncorrelated errors, is an aspect to deal with in future experiments. Also, whether the non-geostrophic currents can be better represented with the DA system and to what extent they are observable by the used observation network is an interesting topic, that has been discussed along the manuscript, but we have not been able to completely answer yet. The velocities represented in the model include both geostrophic and ageostrophic currents. However, most of the kinetic energy contained is due to geostrophic circulation, as it can be seen in Figure 5.1 below (note the difference in the colorbar for the ageostrophic component).

Altimetry observations represent, by nature, only geostrophic circulation. Satellite SST ultra-high products can have resolutions of 1 km, although the effective resolution of the structures it can resolve is about 10km (https://podaac.jpl.nasa.gov/dataset/JPL-L4UHfnd-GLOB-MUR). These SST images can still depict some small scale structures related to ageostrophic circulation but are probably not able to fully resolve them. HF radars, on the other hand, can capture the small scale and high frequency sig-

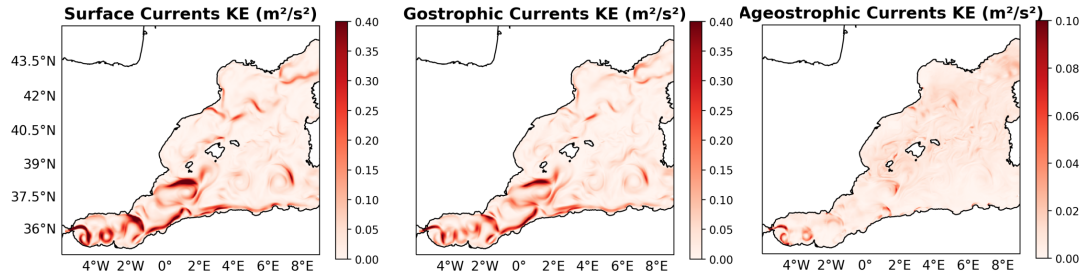

Figure 5.1 *Kinetic energy from the WMOP total Surface currents (left), geostrophic (center), and ageostrophic currents (right) for 16-Oct-2014.*

nals characteristic of non-geostrophic circulation and can be used to correct this kind of processes, especially when using hourly data, as explained in Section 3.1. In our assimilation system, we filter all the high-frequency and focus on correcting the subinertial circulation rather than inertial oscillations. However, in the daily mean field not all the ageostrophic currents are filtered. Nonetheless, we showed in our results, that correcting daily averaged circulation with the assimilation of daily mean HFR observations yields a better representation of some non-geostrophic features in the model. However, we have not specifically evaluated the geostrophic and ageostrophic components of the currents separately. The OSSE could be a good framework to perform further experiments and evaluate this aspect in future works.

Some of these open questions and work-lines will be addressed in the near future. For instance, in EuroSea project where we are engaged, which is an initiative from the European commission with partners all around Europe that intends to improve the European ocean observing and forecasting system in a global context. It is directly linked to the UN decade objectives, and will deliver ocean observations and forecasts to advance scientific knowledge about ocean climate, marine ecosystems and their vulnerability to human impacts. In one of the tasks we are involved in, we would like to evaluate the impact of the assimilation of glider observations in a long-run reanalysis simulation. The glider data from the endurance line that SOCIB gliders continuously realize in the IC will be ingested in the WMOP model, together with glider observations from other data providers, which are distributed in the European CMEMS portal. The objective is to perform a model inter-comparison and evaluate how the continuous assimilation of this data sources could improve the forecasting and the characterization of water mass exchanges and transport processes within sub-basins, before implementing it in the operational chain.

Another upcoming work, within the EuroSea project, which is actually under development is the realization of OSSEs experiments for the support of the future SWOT satellite deployment. The knowledge acquired through this Thesis will certainly be an asset to address this commitment. Different sampling procedures will be assessed to decide the best possible strategy to follow during the calibration phase of the mission. The high-resolution eNATL60 (https://github.com/ocean-next/eNATL60/) simulation developed by Ocean Next and the MEOM group in Grenoble will be used as a nature run to generate pseudo-observations. The outputs of the WMOP data assimilation system will be compared against other field reconstruction methods as Optimal Interpolation or machine learning techniques.

Finally the impact of different HFR systems in the Western Mediterranean, as the

ones from the Ebro Delta or Toulon also needs to be evaluated, providing a potential additional and very valuable contribution to the WMOP system.

Bibliography

Aguiar, E., Mourre, B., Juza, M., Reyes, E., Hernández-Lasheras, J., Cutolo, E., Mason, E., and Tintoré, J. (2020). Multi-platform model assessment in the Western Mediterranean Sea : impact of downscaling on the surface circulation and mesoscale activity. *Ocean Dynamics*.

Alvarez, A. and Mourre, B. (2012). Optimum sampling designs for a glider–mooring observing network. *Journal of Atmospheric and Oceanic Technology*, 29(4):601–612.

Alvarez, A. and Mourre, B. (2014). Cooperation or coordination of underwater glider networks? An assessment from observing system simulation experiments in the Ligurian Sea. *Journal of Atmospheric and Oceanic Technology*, 31(10):2268–2277.

Atlas, R. (1997). Atmospheric observations and experiments to assess their usefulness in data assimilation (gtspecial issueltdata assimilation in meteology and oceanography: theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):111–130.

Aydoğdu, A., Pinardi, N., Pistoia, J., Martinelli, M., Belardinelli, A., and Sparnocchia, S. (2016). Assimilation experiments for the fishery observing system in the Adriatic Sea. *Journal of Marine Systems*, 162:126–136.

Balbín, R., López-Jurado, J. L., Flexas, M. M., Reglero, P., Vélez-Velchí, P., González-Pola, C., Rodríguez, J. M., García, A., and Alemany, F. (2014). Interannual variability of the early summer circulation around the Balearic Islands: Driving factors and potential effects on the marine ecosystem. *Journal of Marine Systems*, 138:70–81.

Ballabrera-Poy, J., Hackert, E., Murtugudde, R., and Busalacchi, A. J. (2007). An observing system simulation experiment for an optimal moored instrument array in the tropical indian ocean. *Journal of Climate*, 20(13):3284–3299.

Barrick, D. E. (2008). 30 Years of CMTC and COPAR. *Proceedings of the IEEE Working Conference on Current Measurement Technology*, pages 131–136.

Barth, A., Alvera-Azcarate, A., Beckers, J.-M., Staneva, J., Stanev, E. V., and Schulz-Stellenfleth, J. (2011). Correcting surface winds by assimilating high-frequency radar surface currents in the German Bight. *Ocean Dynamics*, pages 599–610.

Barth, A., Alvera-Azcárate, A., Troupin, C., Ouberdous, M., and Beckers, J. . (2010). A web interface for griding arbitrarily distributed in situ data based on Data-Interpolating Variational Analysis (DIVA). *Advances in Geosciences*, 28:29.

Barth, A., Alvera-Azcárate, A., and Weisberg, R. H. (2008). Assimilation of high-frequency radar currents in a nested model of the West Florida Shelf. *Journal of Geophysical Research: Oceans*, 113(8):1–15.

Barth, A., Troupin, C., Reyes, E., Alvera-Azcárate, A., Beckers, J.-M., and Tintoré, J. (2021). Variational interpolation of high-frequency radar surface currents using divand. *Ocean Dynamics*, 71(3):293–308.

Baudena, A., Ser-Giacomi, E., D'Onofrio, D., Capet, X., Cotté, C., Cherel, Y., and D'Ovidio, F. (2021). Fine-scale structures as spots of increased fish concentration in the open ocean. *Scientific Reports*, 11(1):1–13.

Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, 53:370–418.

Benkiran, M., Ruggiero, G., Greiner, E., Le Traon, P. Y., Rémy, E., Lellouche, J. M., Bourdallé-Badie, R., Drillet, Y., and Tchonang, B. (2021). Assessing the Impact of the Assimilation of SWOT Observations in a Global High-Resolution Analysis and Forecasting System Part 1: Methods. *Frontiers in Marine Science*, 8(July):1–19.

Bertino, L., Evensen, G., and Wackernagel, H. (2003). Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71(2):223–241.

Bethoux, J. P., Gentili, B., Morin, P., Nicolas, E., Pierre, C., and Ruiz-Pino, D. (1999). The mediterranean sea: a miniature ocean for climatic and environmental studies and a key for the climatic functioning of the north atlantic. *Progress in Oceanography*, 44(1-3):131–146.

Bourlès, B., Lumpkin, R., McPhaden, M. J., Hernandez, F., Nobre, P., Campos, E., Yu, L., Planton, S., Busalacchi, A., Moura, A. D., et al. (2008). The pirata program: History, accomplishments, and future directions. *Bulletin of the American Meteorological Society*, 89(8):1111–1126.

Breivik, O. (2001). Real time assimilation of HF radar currents into a coastal ocean model. *Journal of Marine Systems*, pages 1–25.

Bryan, K. and Cox, M. D. (1967). A numerical investigation of the oceanic general circulation. *Tellus*, 19(1):54–80.

Cabanellas-Reboredo, M., Vazquez-Luis, M., Mourre, B., Alvarez, E., Deudero, S., Amores, A., Addis, P., Ballesteros, E., Barrajon, A., Coppa, S., Garcia-March, J. R., Giacobbe, S., Casalduero, F. G., Hadjioannou, L., Jimenez-Gutierrez, S. V., Katsanevakis, S., Kersting, D., Macic, V., Mavric, B., Patti, F. P., Planes, S., Prado, P., Sanchez, J., Tena-Medialdea, J., de Vaugelas, J., Vicente, N., Belkhamssa, F. Z., Zupan, I., and Hendriks, I. E. (2019). Tracking a mass mortality outbreak of pen shell Pinna nobilis populations: A collaborative effort of scientists and citizens. *Scientific Reports*, 9(1):1–11.

Calò, A., Lett, C., Mourre, B., Perez-Ruzafa, A., and García-Charton, J. A. (2018). Use of Lagrangian simulations to hindcast the geographical position of propagule release zones in a Mediterranean coastal fish. *Marine Environmental Research*, 134(September):16–27.

Capet, A., Fernández, V., She, J., Dabrowski, T., Umgiesser, G., Staneva, J., Mészáros, L., Campuzano, F., Ursella, L., Nolan, G., et al. (2020). Operational modeling capacity in european seas—an eurogoos perspective and recommendations for improvement. *Frontiers in Marine Science*, 7:129.

Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):1–50.

Chapman, R. D., Shay, L. K., Graber, H. C., Edson, J. B., Karachintsev, A., Trump, C. L., and Ross, D. B. (1997). On the accuracy of HF radar surface current measurements: Intercomparisons with ship-based sensors. *Journal of Geophysical Research: Oceans*, 102(C8):18737–18748.

Chassignet, E. P. (2021). Ocean modeling. In *Preparing a Workforce for the New Blue Economy*, chapter 4, pages 47–61. Elsevier.

Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., Baraille, R., and Bleck, R. (2007). The hycom (hybrid coordinate ocean model) data assimilative system. *Journal of Marine Systems*, 65(1-4):60–83.

Chelton, D. B., Schlax, M. G., and Samelson, R. M. (2011). Global observations of nonlinear mesoscale eddies. *Progress in oceanography*, 91(2):167–216.

Chelton, D. B., Schlax, M. G., Samelson, R. M., and de Szoeke, R. A. (2007). Global observations of large oceanic eddies. *Geophysical Research Letters*, 34(15).

Chin, T. M., Vazquez-Cuervo, J., and Armstrong, E. M. (2017). A multi-scale high-resolution analysis of global sea surface temperature. *Remote Sensing of Environment*, 200(June):154–169.

CIESIN, Center for International Earth Science Information Network. (2021). National Aggregates of Geospatial Data: Population, Landscape and Climate Estimates Version 3.

Compa, M., Alomar, C., Mourre, B., March, D., Tintoré, J., and Deudero, S. (2020). Nearshore spatio-temporal sea surface trawls of plastic debris in the Balearic Islands. *Marine Environmental Research*, 158:104945.

Counillon, F. and Bertino, L. (2009). Ensemble Optimal Interpolation: Multivariate properties in the Gulf of Mexico. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 61(2):296–308.

Cózar, A., Sanz-Martín, M., Martí, E., González-Gordillo, J. I., Ubeda, B., Gálvez, J. Á., Irigoien, X., and Duarte, C. M. (2015). Plastic accumulation in the mediterranean sea. *PloS one*, 10(4):e0121762.

Crombie, D. (1971). Backscatter of hf radio waves from the sea. *Electromagnetic Probing in Geophysics*, 79:131.

Crombie, D. D. (1955). Doppler Spectrum of Sea Echo at 13.56 Mc./s. *Nature*, 175(4459):681–682.

Davidson, F., Alvera-Azcárate, A., Barth, A., Brassington, G. B., Chassignet, E. P., Clementi, E., De Mey-Frémaux, P., Divakaran, P., Harris, C., Hernandez, F., Hogan, P., Hole, L. R., Holt, J., Liu, G., Lu, Y., Lorente, P., Maksymczuk, J., Martin, M., Mehra, A., Melsom, A., Mo, H., Moore, A., Oddo, P., Pascual, A., Pequignet, A. C.,

Kourafalou, V., Ryan, A., Siddorn, J., Smith, G., Spindler, D., Spindler, T., Stanev, E. V., Staneva, J., Storto, A., Tanajura, C., Vinayachandran, P. N., Wan, L., Wang, H., Zhang, Y., Zhu, X., and Zu, Z. (2019). Synergies in operational oceanography: The intrinsic need for sustained ocean observations. *Frontiers in Marine Science*, 6(JUL):1–18.

De Mey-Frémaux, P., Ayoub, N., Barth, A., Brewin, R., Charria, G., Campuzano, F., Ciavatta, S., Cirano, M., Edwards, C. A., Federico, I., Gao, S., Garcia Hermosa, I., Garcia Sotillo, M., Hewitt, H., Hole, L. R., Holt, J., King, R., Kourafalou, V., Lu, Y., Mourre, B., Pascual, A., Staneva, J., Stanev, E. V., Wang, H., and Zhu, X. (2019). Model-Observations Synergy in the Coastal Ocean. *Frontiers in Marine Science*, 6.

Delaney, J. and Barga, R. (2009). Vision for ocean science. *Hey T, Tansley S, Tolle K. The Fourth Paradigm. Washington: Microsoft Research*.

deYoung, B., Visbeck, M., Filho, M. C., Baringer, M. O., Black, C. A., Buch, E., Canonico, G., Coelho, P., Duha, J. T., Edwards, M., Fischer, A. S., Fritz, J. S., Ketelhake, S., Muelbert, J. H., Monteiro, P., Nolan, G., O'Rourke, E., Ott, M., Le Traon, P. Y., Pouliquen, S., PInto, I. S., Tanhua, T., Velho, F., and Willis, Z. (2019). An integrated all-Atlantic ocean observing system in 2030. *Frontiers in Marine Science*, 6(JUL):1–22.

Dobricic, S., Pinardi, N., Testor, P., and Send, U. (2010). Impact of data assimilation of glider observations in the Ionian Sea (Eastern Mediterranean). *Dynamics of Atmospheres and Oceans*, 50(1):78–92.

d'Ovidio, F., Fernández, V., Hernández-García, E., and López, C. (2004). Mixing structures in the Mediterranean Sea from finite-size Lyapunov exponents. *Geophysical Research Letters*, 31(17):1–4.

Emery, B. M., Washburn, L., and Harlan, J. A. (2004). Evaluating radial current measurements from CODAR high-frequency radars with moored current meters. *Journal of Atmospheric and Oceanic Technology*, 21(8):1259–1271.

Escudier, R., Mourre, B., Juza, M., and Tintoré, J. (2016a). Subsurface circulation and mesoscale variability in the Algerian subbasin from altimeter-derived eddy trajectories. *Journal of Geophysical Research: Oceans*, 121(8):6310–6322.

Escudier, R., Renault, L., Pascual, A., Brasseur, P., Chelton, D., and Beuvier, J. (2016b). Eddy properties in the western mediterranean sea from satellite altimetry and a numerical simulation. *Journal of Geophysical Research: Oceans*, 121(6):3990–4006.

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367.

Farcy, P., Durand, D., Charria, G., Painting, S. J., Tamminem, T., Collingridge, K., Grémare, A. J., Delauney, L., and Puillat, I. (2019a). Toward a European coastal observing network to provide better answers to science and to societal challenges; the JERICO research infrastructure. *Frontiers in Marine Science*, 6(SEP):1–13.

Farcy, P., Durand, D., Charria, G., Painting, S. J., Tamminem, T., Collingridge, K., Grémare, A. J., Delauney, L., and Puillat, I. (2019b). Toward a European coastal observing network to provide better answers to science and to societal challenges; the JERICO research infrastructure. *Frontiers in Marine Science*, 6(SEP):1–13.

Fox-Kemper, B., Adcroft, A., Böning, C. W., Chassignet, E. P., Curchitser, E., Danabasoglu, G., Eden, C., England, M. H., Gerdes, R., Greatbatch, R. J., Griffies, S. M., Hallberg, R. W., Hanert, E., Heimbach, P., Hewitt, H. T., Hill, C. N., Komuro, Y., Legg, S., Sommer, J. L., Masina, S., Marsland, S. J., Penny, S. G., Qiao, F., Ringler, T. D., Treguier, A. M., Tsujino, H., Uotila, P., and Yeager, S. G. (2019). Challenges and prospects in ocean circulation models. *Frontiers in Marine Science*, 6(FEB):1–29.

Fujii, Y., Remy, E., Zuo, H., Oke, P. R., Halliwell, G. R., Gasparin, F., Benkiran, M., Loose, N., Cummings, J., Xie, J., Xue, Y., Masuda, S., Smith, G. C., Balmaseda, M. A., Germineaud, C., Lea, D. J., Larnicol, G., Bertino, L., Bonaduce, A., Brasseur, P., Donlon, C., Heimbach, P., Kim, Y. H., Kourafalou, V., Le Traon, P. Y., Martin, M. J., Paturi, S., Tranchant, B., and Usui, N. (2019). Observing system evaluation based on ocean data assimilation and prediction systems: On-going challenges and future vision for designing/supporting ocean observational networks. *Frontiers in Marine Science*, 6(JUL):1–25.

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

Gasparin, F., Guinehut, S., Mao, C., Mirouze, I., Rémy, E., King, R. R., Hamon, M., Reid, R., Storto, A., Le Traon, P.-Y., Martin, M. J., and Masina, S. (2019). Requirements for an Integrated in situ Atlantic Ocean Observing System From Coordinated Observing System Simulation Experiments. *Frontiers in Marine Science*, 6(February).

Guinehut, S., Dhomps, A. L., Larnicol, G., and Le Traon, P. Y. (2012). High resolution 3-D temperature and salinity fields derived from in situ and satellite observations. *Ocean Science*, 8(5):845–857.

Guinehut, S., Le Traon, P. Y., Larnicol, G., and Philipps, S. (2004). Combining Argo and remote-sensing data to estimate the ocean three-dimensional temperature fields - A first approach based on simulated observations. *Journal of Marine Systems*, 46(1-4):85–98.

Halliwell, G. R., Goni, G. J., Mehari, M. F., Kourafalou, V. H., Baringer, M., and Atlas, R. (2020). OSSE Assessment of Underwater Glider Arrays to Improve Ocean Model Initialization for Tropical Cyclone Prediction. *Journal of Atmospheric and Oceanic Technology*, 37(3):467–487.

Halliwell, J. R., Srinivasan, A., Kourafalou, V., Yang, H., Willey, D., Le Hénaff, M., and Atlas, R. (2014). Rigorous evaluation of a fraternal twin ocean OSSE system for the open Gulf of Mexico. *Journal of Atmospheric and Oceanic Technology*, 31(1):105–130.

Hayes, D. R., Dobricic, S., Gildor, H., and Matsikaris, A. (2019). Operational assimilation of glider temperature and salinity for an improved description of the Cyprus eddy. *Deep Sea Research Part II: Topical Studies in Oceanography*, 164:41–53.

Hernández-Carrasco, I., López, C., Hernández-García, E., and Turiel, A. (2011). How reliable are finite-size Lyapunov exponents for the assessment of ocean dynamics? *Ocean Modelling*, 36(3-4):208–218.

Hernandez-Carrasco, I., López, C., Hernandez-García, E., and Turiel, A. (2012). Seasonal and regional characterization of horizontal stirring in the global ocean. *Journal of Geophysical Research: Oceans*, 117(10):1–12.

Hernández-Carrasco, I., Orfila, A., Rossi, V., and Garçon, V. (2018). Effect of small scale transport processes on phytoplankton distribution in coastal seas. *Scientific Reports*, 8(1):1–13.

Hernández-Carrasco, I., Rossi, V., Hernández-García, E., Garçon, V., and López, C. (2014). The reduction of plankton biomass induced by mesoscale stirring: A modeling study in the benguela upwelling. *Deep Sea Research Part I: Oceanographic Research Papers*, 83:65–80.

Hernandez-Lasheras, J. and Mourre, B. (2018). Dense CTD survey versus glider fleet sampling: comparison of the performance for regional ocean prediction West of Sardinia. *Ocean Science*, 14:1069–1084.

Hernandez-Lasheras, J., Mourre, B., Orfila, A., Santana, A., Reyes, E., and Tintoré, J. (2021). Evaluating High-Frequency radar data assimilation impact in coastal ocean operational modelling. *Ocean Science Discussions*, 17:1157–1175.

Heslop, E. E., Ruiz, S., Allen, J., López-jurado, J. L., Renault, L., and Tintoré, J. (2012). Autonomous underwater gliders monitoring variability at "choke points" in our ocean system: A case study in the Western Mediterranean Sea. *Geophysical Research Letters*, 39:1–6.

Hoffman, R. N. and Atlas, R. (2016). Future observing system simulation experiments. *Bulletin of the American Meteorological Society*, 97(9):1601–1616.

Hoteit, I., Luo, X., Bocquet, M., Kohl, A., and Ait-El-Fquih, B. (2018). Data Assimilation in Oceanography: Current Status and New Directions. *New Frontiers in Operational Oceanography*, pages 465–511.

Iermano, I., Moore, A. M., and Zambianchi, E. (2016a). Impacts of a 4-dimensional variational data assimilation in a coastal ocean model of southern Tyrrhenian Sea. *Journal of Marine Systems*, 154:157–171.

Iermano, I., Moore, A. M., and Zambianchi, E. (2016b). Impacts of a 4-dimensional variational data assimilation in a coastal ocean model of southern Tyrrhenian Sea. *Journal of Marine Systems*, 154:157–171.

International-Altimetry-Team (2021). Altimetry for the future: Building on 25 years of progress. *Advances in Space Research*, 68:319–363.

Janeković, I., Mihanović, H., Vilibić, I., Grčić, B., Ivatek-Šahdan, S., Tudor, M., and Djakovac, T. (2020a). Using multi-platform 4D-Var data assimilation to improve modeling of Adriatic Sea dynamics. *Ocean Modelling*, 146(November 2019):101538.

Janeković, I., Mihanović, H., Vilibić, I., Grčić, B., Ivatek-Šahdan, S., Tudor, M., and Djakovac, T. (2020b). Using multi-platform 4D-Var data assimilation to improve modeling of Adriatic Sea dynamics. *Ocean Modelling*, 146:101538.

Juza, M., Mourre, B., Renault, L., Gómara, S., Sebastian, K., López, S. L., Borrueco, B. F., Beltran, J., Troupin, C., Tomás, M. T., et al. (2016). Operational SOCIB forecasting system and multi-platform validation in the Western Mediterranean. *Jounal of Operational Oceanography*, 9:9231.

Kai, E. T., Rossi, V., Sudre, J., Weimerskirch, H., Lopez, C., Hernandez-Garcia, E., Marsac, F., and Garçon, V. (2009). Top marine predators track lagrangian coherent structures. *Proceedings of the National Academy of Sciences*, 106(20):8245–8250.

Kaplan, D. M. and Lekien, F. (2007). Spatial interpolation and filtering of surface current data based on open-boundary modal analysis. *Journal of Geophysical Research: Oceans*, 112(12):1–20.

Kerry, C., Powell, B., Roughan, M., and Oke, P. (2016). Development and evaluation of a high-resolution reanalysis of the East Australian Current region using the Regional Ocean Modelling System (ROMS 3.4) and Incremental Strong-Constraint 4-Dimensional Variational (IS4D-Var) data assimilation. *Geoscientific Model Development*, 9(10):3779–3801.

Kerry, C., Roughan, M., and Powell, B. (2018). Observation Impact in a Regional Reanalysis of the East Australian Current System. *Journal of Geophysical Research: Oceans*, pages 1–18.

Kersting, D. K., Vázquez-Luis, M., Mourre, B., Belkhamssa, F. Z., Álvarez, E., Bakran-Petricioli, T., Barberá, C., Barrajón, A., Cortés, E., Deudero, S., García-March, J. R., Giacobbe, S., Giménez-Casalduero, F., González, L., Jiménez-Gutiérrez, S., Kipson, S., Llorente, J., Moreno, D., Prado, P., Pujol, J. A., Sánchez, J., Spinelli, A., Valencia, J. M., Vicente, N., and Hendriks, I. E. (2020). Recruitment Disruption and the Role of Unaffected Populations for Potential Recovery After the Pinna nobilis Mass Mortality Event. *Frontiers in Marine Science*, 7(October):1–11.

Knoll, M., Borrione, I., Fiekas, H.-V., Funk, A., Hemming, M. P., Kaiser, J., Onken, R., Queste, B., and Russo, A. (2017). Hydrography and circulation west of sardinia in june 2014. *Ocean Science*, 13(6):889–904.

Korres, G., Ntoumas, M., Potiris, M., and Petihakis, G. (2014). Assimilating ferry box data into the Aegean Sea model. *Journal of Marine Systems*, 140:59–72.

Kourafalou, V., De Mey, P., Le Hénaff, M., Charria, G., Edwards, C., He, R., Herzfeld, M., Pascual, a., Stanev, E., Tintoré, J., Usui, N., van der Westhuysen, a., Wilkin, J., and Zhu, X. (2015a). Coastal Ocean Forecasting: system integration and evaluation. *Journal of Operational Oceanography*, 8(sup1):s127–s146.

Kourafalou, V., De Mey, P., Staneva, J., Ayoub, N., Barth, a., Chao, Y., Cirano, M., Fiechter, J., Herzfeld, M., Kurapov, a., Moore, a., Oddo, P., Pullen, J., van der Westhuysen, a., and Weisberg, R. (2015b). Coastal Ocean Forecasting: science foundation and user benefits. *Journal of Operational Oceanography*, 8(sup1):s147–s167.

Lana, A., Fernández, V., and Tintoré, J. (2015). SOCIB Continuous Observations Of Ibiza Channel Using HF Radar. *Sea Technology 56*.

Lana, A., Marmain, J., Fernandez, V., Tintore, J., and Orfila, A. (2016). Wind influence on surface current variability in the Ibiza Channel from HF Radar. *Ocean Dynamics*, pages 483–497.

Lange, M. and Van Sebille, E. (2017). Parcels v0.9: Prototyping a Lagrangian Ocean Analysis framework for the petascale age. *arXiv*, pages 4175–4186.

Le Sommer, J., Chassignet, E. P., and Wallcraft, A. J. (2018). Ocean Circulation Modeling for Operational Oceanography: Current Status and Future Challenges. *New Frontiers in Operational Oceanography*, pages 289–306.

Le Traon, P. Y. (2013). From satellite altimetry to Argo and operational oceanography: Three revolutions in oceanography. *Ocean Science*, 9(5):901–915.

Le Traon, P. Y., Reppucci, A., Alvarez Fanjul, E., Aouf, L., Behrens, A., and Belmonte, M. e. a. (2019). From Observation to Information and Users: The Copernicus Marine Service Perspective. *Frontiers in Marine Science*, 6.

Lee, T. and Gentemann, C. (2018). Satellite SST and SSS Observations and Their Role to Constrain Ocean Models. *New Frontiers in Operational Oceanography*, pages 271–288.

Lehahn, Y., d'Ovidio, F., Lévy, M., and Heifetz, E. (2007). Stirring of the northeast Atlantic spring bloom: A Lagrangian analysis based on multisatellite data. *Journal of Geophysical Research: Oceans*, 112(8):1–15.

Lekien, F., Coulliette, C., Mariano, A. J., Ryan, E. H., Shay, L. K., Haller, G., and Marsden, J. (2005). Pollution release tied to invariant manifolds: A case study for the coast of florida. *Physica D: Nonlinear Phenomena*, 210(1-2):1–20.

Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R., Garric, G., Melet, A., Drévillon, M., Bricaud, C., Hamon, M., Le Galloudec, O., Regnier, C., et al. (2021). The copernicus global 1/12° oceanic and sea ice glorys12 reanalysis. *Frontiers in Earth Science*, 9.

Lellouche, J. M., Greiner, E., Le Galloudec, O., Garric, G., Regnier, C., Drevillon, M., Benkiran, M., Testut, C. E., Bourdalle-Badie, R., Gasparin, F., Hernandez, O., Levier, B., Drillet, Y., Remy, E., and Le Traon, P. Y. (2018). Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1g 12° high-resolution system. *Ocean Science*, 14(5):1093–1126.

Lellouche, J.-M., Le Galloudec, O., Drévillon, M., Régnier, C., Greiner, E., Garric, G., Ferry, N., Desportes, C., Testut, C.-E., Bricaud, C., et al. (2013). Evaluation of global monitoring and forecasting systems at mercator océan. *Ocean Science*, 9(1):57–81.

Lermusiaux, P. F. (2007). Adaptive modeling, adaptive data assimilation and adaptive sampling. *Physica D: Nonlinear Phenomena*, 230(1-2):172–196.

Li, Z., McWilliams, J. C., Ide, K., and Farrara, J. D. (2015). Coastal ocean data assimilation using a multi-scale three-dimensional variational scheme. *Ocean Dynamics*, 65(7):1001–1015.

Liblik, T., Karstensen, J., Testor, P., Alenius, P., Hayes, D., Ruiz, S., Heywood, K. J., Pouliquen, S., Mortier, L., and Mauri, E. (2016). Potential for an underwater glider component as part of the Global Ocean Observing System. *Methods in Oceanography*, 17:50–82.

Lipa, B. J. and Barrick, D. E. (1983). Least-squares methods for the extraction of surface currents from CODAR cross-loop data: Application at ARSLOE. *IEEE Transactions on Geoscience and Remote Sensing*, 8(4):226–253.

Liu, Y. and Weisberg, R. H. (2011). Evaluation of trajectory modeling in different dynamic regions using normalized cumulative Lagrangian separation. *Journal of Geophysical Research: Oceans*, 116(9):1–13.

Lorente, P., Lin-Ye, J., García-León, M., Reyes, E., Fernandes, M., Sotillo, M. G., Espino, M., Ruiz, M. I., Gracia, V., Perez, S., Aznar, R., Alonso-Martirena, A., and Álvarez-Fanjul, E. (2021a). On the Performance of High Frequency Radar in the Western Mediterranean During the Record-Breaking Storm Gloria.

Lorente, P., Lin-Ye, J., García-León, M., Reyes, E., Fernandes, M., Sotillo, M. G., Espino, M., Ruiz, M. I., Gracia, V., Perez, S., et al. (2021b). On the performance of high frequency radar in the western mediterranean during the record-breaking storm gloria. *Frontiers in Marine Science*, 8:205.

Lorente, P., Piedracoba, S., Soto-Navarro, J., and Alvarez-Fanjul, E. (2015). Evaluating the surface circulation in the Ebro delta (northeastern Spain) with quality-controlled high-frequency radar measurements. *Ocean Science*, 11(6):921–935.

Malanotte-Rizzoli, P., Artale, V., Borzelli-Eusebi, G., Brenner, S., Crise, A., Gacic, M., Kress, N., Marullo, S., Ribera d'Alcalà, M., Sofianos, S., et al. (2014). Physical forcing and physical/biochemical variability of the mediterranean sea: a review of unresolved issues and directions for future research. *Ocean Science*, 10(3):281–322.

Mantovani, C., Corgnati, L., Horstmann, J., Rubio, A., Reyes, E., Quentin, C., Cosoli, S., Asensio, J. L., Mader, J., and Griffa, A. (2020). Best Practices on High Frequency Radar Deployment and Operation for Ocean Current Measurement. *Frontiers in Marine Science*, 7(April):1–21.

Marmain, J., Molcard, A., Forget, P., Barth, A., and Toulon, D. (2014). Assimilation of HF radar surface currents to optimize forcing in the northwestern Mediterranean Sea. *Nonlinear Processes in Geophysics*, pages 659–675.

Martin, M. J., Balmaseda, M., Bertino, L., Brasseur, P., Brassington, G., Cummings, J., Fujii, Y., Lea, D. J., Lellouche, J. M., Mogensen, K., Oke, P. R., Smith, G. C., Testut, C. E., Waagbø, G. A., Waters, J., and Weaver, A. T. (2015). Status and future of data assimilation in operational oceanography. *Journal of Operational Oceanography*, 8(0):s28–s48.

Mason, E., Pascual, A., and McWilliams, J. C. (2014). A new sea surface height–based code for oceanic mesoscale eddy tracking. *Journal of Atmospheric and Oceanic Technology*, 31(5):1181–1188.

McWilliams, J. C. (1996). Modeling the oceanic general circulation. *Annual Review of Fluid Mechanics*, 28(1):215–248.

Melet, A., Verron, J., and Brankart, J. (2012). Potential outcomes of glider data assimilation in the Solomon Sea: Control of the water mass properties and parameter estimation. *Journal of Marine Systems*, 94:232–246.

Millot, C. (1999). Circulation in the Western Mediterranean Sea. *Journal of Marine Systems*, 20(1-4):423–442.

Millot, C. and Taupier-Letage, I. (2005). Circulation in the mediterranean sea. In *The Mediterranean Sea*, pages 29–66. Springer.

Moltmann, T., Turton, J., Zhang, H. M., Nolan, G., Gouldman, C., Griesbauer, L., Willis, Z., Piniella, A. M., Barrell, S., Andersson, E., Gallage, C., Charpentier, E., Belbeoch, M., Poli, P., Rea, A., Burger, E. F., Legler, D. M., Lumpkin, R., Meinig, C., O'Brien, K., Saha, K., Sutton, A., Zhang, D., and Zhang, Y. (2019). A Global Ocean Observing System (GOOS), delivered through enhanced collaboration across regions, communities, and new technologies. *Frontiers in Marine Science*, 6(JUN):1–21.

Moore, A. M., Martin, M. J., Akella, S., Arango, H. G., Balmaseda, M., Bertino, L., Ciavatta, S., Cornuelle, B., Cummings, J., Frolov, S., Lermusiaux, P., Oddo, P., Oke, P. R., Storto, A., Teruzzi, A., Vidard, A., and Weaver, A. T. (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(March):1–6.

Morrow, R., Blurmstein, D., and Dibarboure, G. (2018). Fine-scale Altimetry and the Future SWOT Mission. *New Frontiers in Operational Oceanography*, pages 191–226.

Mourre, B., Aguiar, E., Juza, M., Hernandez-Lasheras, J., Reyes, E., Heslop, E., Escudier, R., Cutolo, E., Ruiz, S., Mason, E., Pascual, A., and Tintoré, J. (2018). Assessment of High-Resolution Regional Ocean Prediction Systems Using Multi-Platform Observations: Illustrations in the Western Mediterranean Sea. *New Frontiers in Operational Oceanography*, pages 663–694.

Mourre, B. and Alvarez, A. (2012). Benefit assessment of glider adaptive sampling in the Ligurian Sea. *Deep-Sea Research Part I: Oceanographic Research Papers*, 68:68–78.

Mourre, B. and Chiggiato, J. (2014). A comparison of the performance of the 3-D super-ensemble and an ensemble Kalman filter for short-range regional ocean prediction. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 66(1):21640.

Mourre, B., Crosnier, L., and Provost, C. L. (2006a). Real-time sea-level gauge observations and operational oceanography. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 364(1841):867–884.

Mourre, B., De Mey, P., Ménard, Y., Lyard, F., and Le Provost, C. (2006b). Relative performance of future altimeter systems and tide gauges in constraining a model of north sea high-frequency barotropic dynamics. *Ocean Dynamics*, 56(5):473–486.

NASA/JPL (2015). GHRSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1).

Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D., and Kosro, P. M. (2002). Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research C: Oceans*, 107(9):5–1.

Oke, P. R., Balmaseda, M. A., Benkiran, M., Cummings, J. A., Dombrowsky, E., Fujii, Y., Guinehut, S., Larnicol, G., Le Traon, P.-Y., and Martin, M. J. (2009). Observing system evaluations: Using godae systems. *Oceanography*, 22(3):144–153.

Oke, P. R., Brassington, G. B., Griffin, D. A., and Schiller, A. (2008). The Bluelink ocean data assimilation system (BODAS). *Ocean modelling*, 21:46–70.

Oke, P. R., Larnicol, G., Fujii, Y., Smith, G. C., Lea, D. J., Guinehut, S., Remy, E., Balmaseda, M. A., Rykova, T., Surcel-Colan, D., et al. (2015a). Assessing the impact of observations on ocean forecasts and reanalyses: Part 1, global studies. *Journal of Operational Oceanography*, 8(sup1):s49–s62.

Oke, P. R., Larnicol, G., Jones, E. M., Kourafalou, V., Sperrevik, A., Carse, F., Tanajura, C. A., Mourre, B., Tonani, M., Brassington, G., et al. (2015b). Assessing the impact of observations on ocean forecasts and reanalyses: Part 2, regional applications. *Journal of Operational Oceanography*, 8(sup1):s63–s79.

Oke, P. R. and Sakov, P. (2008). Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology*, 25(6):1004–1017.

Oke, P. R., Sakov, P., and Corney, S. P. (2007). Impacts of localisation in the EnKF and EnOI: Experiments with a small model. *Ocean Dynamics*, 57(1):32–45.

Olita, A., Ribotti, A., Fazioli, L., Perilli, A., and Sorgente, R. (2013). Surface circulation and upwelling in the Sardinia Sea_ A numerical study. *Continental Shelf Research*, 71:95–108.

Olita, A., Ribotti, A., Sorgente, R., Fazioli, L., and Perilli, A. (2011). SLA–chlorophyll-a variability and covariability in the Algero-Provençal Basin (1997–2007) through combined use of EOF and wavelet analysis of satellite data. *Ocean Dynamics*, 61:89–102.

Onken, R., Álvarez, A., Fernández, V., Vizoso, G., Basterretxea, G., Tintoré, J., Haley, P., and Nacini, E. (2008). A forecast experiment in the Balearic Sea. *Journal of Marine Systems*, 71(1-2):79–98.

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

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, 56(5):415–428.

Paduan, J. D., Kim, K. C., Cook, M. S., and Chavez, F. P. (2006). Calibration and validation of direction-finding high-frequency radar ocean surface current observations. *IEEE Journal of Oceanic Engineering*, 31(4):862–875.

Paduan, J. D. and Shulman, I. (2004). HF radar data assimilation in the Monterey Bay area. *Journal of Geophysical Research*, 109:1–17.

Paduan, J. D. and Washburn, L. (2013). High-Frequency Radar Observations of Ocean Surface Currents. *Annual Review of Marine Science*, 5(1):115–136.

Pan, C., Zheng, L., Weisberg, R. H., Liu, Y., and Lembke, C. E. (2014). Comparisons of different ensemble schemes for glider data assimilation on West Florida Shelf. *OCEAN MODELLING*, 81:13–24.

Pascual, A., Bouffard, J., Ruiz, S., Buongiorno Nardelli, B., Vidal-Vijande, E., Escudier, R., Sayol, J. M., and Orfila, A. (2013). Recent improvements in mesoscale characterization of the western Mediterranean Sea: synergy between satellite altimetry and other observational approaches. *Scientia Marina*, 77(1 SE - Articles):19–36.

Pascual, A., Lana, A., Troupin, C., Ruiz, S., Faugère, Y., Escudier, R., and Tintoré, J. (2015). Assessing SARAL/AltiKa Data in the Coastal Zone: Comparisons with HF Radar Observations. *Marine Geodesy*, 38:260–276.

Pascual, A. and Macías, D. (2021). *Ocean Science Challenges for 2030*, volume 13. CSIC.

Pascual, A., Ruiz, S., Olita, A., Troupin, C., Claret, M., Casas, B., Mourre, B., Poulain, P.-M., Tovar-sanchez, A., Capet, A., Mason, E., Allen, J. T., Mahadevan, A., and Tintore, J. (2017). A multiplatform experiment to unravel meso- and submesoscale processes in an intense front (AlborEx). *Frontiers in Marine Science*.

Pearson, B. and Fox-Kemper, B. (2018). Log-normal turbulence dissipation in global ocean models. *Physical review letters*, 120(9):094501.

Piñeiro, S. (2021). *Thermohaline variability and mixing dynamics in the Western Mediterranean deep waters within the Western Mediterranean transition*. PhD thesis, Universitat de les Illes Balears.

Pinot, J. M., López-Jurado, J. L., and Riera, M. (2002). The CANALES experiment (1996-1998). Interannual, seasonal, and mesoscale variability of the circulation in the Balearic Channels. *Progress in Oceanography*, 55(3-4):335–370.

Pinot, J. M., Tintoré, J., and Gomis, D. (1994). Quasi-synoptic mesoscale variability in the Balearic Sea. *Deep-Sea Research Part I*, 41(5-6):897–914.

Pinot, J. M., Tintoré, J., and Gomis, D. (1995). Multivariate analysis of the surface circulation in the Balearic Sea. *Progress in Oceanography*, 36(4):343–376.

Ren, L., Nash, S., and Hartnett, M. (2016). Forecasting of Surface Currents via Correcting Wind Stress with Assimilation of High-Frequency Radar Data in a Three-Dimensional Model. *Advances in Meteorology*, 2016.

Révelard, A., Reyes, E., Mourre, B., Hernández-Carrasco, I., Rubio, A., Lorente, P., Fernández, C. D. L., Mader, J., Álvarez-Fanjul, E., and Tintoré, J. (2021). Sensitivity of Skill Score Metric to Validate Lagrangian Simulations in Coastal Areas: Recommendations for Search and Rescue Applications.

Ribotti, A., Puillat, I., Sorgente, R., and Natale, S. (2004). Mesoscale circulation in the surface layer off the southern and western Sardinia Island in 2000-2002. *Chemistry and Ecology*, 20(5):345–363.

Rio, M.-H., Pascual, A., Poulain, P.-M., Menna, M., Barceló, B., and Tintoré, J. (2014). Computation of a new mean dynamic topography for the Mediterranean Sea from model outputs, altimeter measurements and oceanographic in situ data. *Ocean Science*, 10(4):731–744.

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. (2019). The Global High Frequency Radar Network. *Frontiers in Marine Science*, 6(May).

Robinson, R. A., Leslie, G. W., Theocharis, A., and Lascaratos, A. (2001). Mediterranean Sea Circulation. *Ocean Currents*, pages 1–19.

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., Oke, P., Owens, W. B., Pouliquen, S., Purkey, S., et al. (2019). On the future of argo: A global, full-depth, multi-disciplinary array. *Frontiers in Marine Science*, 6:439.

Roesler, C. J., Emery, W. J., and Kim, S. Y. (2013). Evaluating the use of high-frequency radar coastal currents to correct satellite altimetry. *Journal of Geophysical Research: Oceans*, 118(7):3240–3259.

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. (2017a). HF Radar activity in European coastal seas: Next steps toward a Pan-European HF Radar network.

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. (2017b). HF Radar activity in European coastal seas: Next steps toward a Pan-European HF Radar network.

Rudnick, D. L. (2016). Ocean Research Enabled by Underwater Gliders. *Annual Review of Marine Science*, 8(1):519–541.

Ruiz, S., Garau, B., Martínez-ledesma, M., Casas, B., Pascual, A., Vizoso, G., Bouffard, J., Heslop, E., Alvarez, A., Testor, P., and Tintoré, J. (2012). New technologies for marine research: five years of glider activities at IMEDEA; Nuevas tecnologías para la investigación marina: 5 años de actividades de gliders en el IMEDEA. *Advances in Spanish physical Oceanography*.

Ruiz-Orejón, L. F., Mourre, B., Sardá, R., Tintoré, J., and Ramis-Pujol, J. (2019). Quarterly variability of floating plastic debris in the marine protected area of the Menorca Channel (Spain). *Environmental Pollution*, 252:1742–1754.

Ryabinin, V., Barbière, J., Haugan, P., Kullenberg, G., Smith, N., McLean, C., Troisi, A., Fischer, A. S., Aricò, S., Aarup, T., Pissierssens, P., Visbeck, M., Enevoldsen, H., and Rigaud, J. (2019). The UN decade of ocean science for sustainable development. *Frontiers in Marine Science*, 6(JUL).

Sakov, P. and Oke, P. R. (2008). Objective array design: Application to the tropical indian ocean. *Journal of atmospheric and oceanic technology*, 25(5):794–807.

Sakov, P. and Sandery, P. A. (2015). Comparison of EnOI and EnKF regional ocean reanalysis systems. *Ocean Modelling*, 89(February 2015):45–60.

Sánchez-Román, A., Ruiz, S., Pascual, A., Mourre, B., and Guinehut, S. (2017). On the mesoscale monitoring capability of Argo floats in the Mediterranean Sea. *Ocean Science*, 13(2):223–234.

Sandery, P. A., Brassington, G. B., and Freeman, J. (2011). Adaptive nonlinear dynamical initialization. *Journal of Geophysical Research*, 116(C1):C01021.

Santinelli, C., Ribotti, A., Sorgente, R., Gasparini, G. P., Nannicini, L., Vignudelli, S., and Seritti, A. (2008). Coastal dynamics and dissolved organic carbon in the western Sardinian shelf (Western Mediterranean). *Journal of Marine Systems*, 74(1-2):167–188.

Schiller, A., Bell, M., Brassington, G., Brasseur, P., Barciela, R., De Mey, P., Dombrowsky, E., Gehlen, M., Hernandez, F., Kourafalou, V., Larnicol, G., Le Traon, P.-Y., Martin, M., Oke, P., Smith, G. C., Smith, N., Tolman, H., and Wilmer-Becker, K. (2015). Synthesis of new scientific challenges for GODAE OceanView. *Journal of Operational Oceanography*, 8(sup2):s259–s271.

Schiller, A., Mourre, B., Drillet, Y., and Brassington, G. (2018). An Overview of Operational Oceanography. *New Frontiers in Operational Oceanography*, pages 1–26.

Shadden, S. C., Lekien, F., and Marsden, J. E. (2005). Definition and properties of lagrangian coherent structures from finite-time lyapunov exponents in two-dimensional aperiodic flows. *Physica D: Nonlinear Phenomena*, 212(3-4):271–304.

Shchepetkin, A. F. and McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, 9(4):347–404.

Shulman, I. and Paduan, J. D. (2009). Assimilation of HF radar-derived radials and total currents in the Monterey Bay area. *Deep-Sea Research II journal*, 56:149–160.

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

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

Sloyan, B. M., Wilkin, J., Hill, K. L., Chidichimo, M. P., Cronin, M. F., Johannessen, J. A., Karstensen, J., Krug, M., Lee, T., Oka, E., Palmer, M. D., Rabe, B., Speich, S., Von Schuckmann, K., Weller, R. A., and Yu, W. (2019). Evolving the global ocean observing system for research and application services through international coordinatio. *Frontiers in Marine Science*, 6(JUL).

Smith, W. H. and Sandwell, D. T. (1997). Global Sea Floor Topography from Satellite Altimetry and Ship Depth Soundings. *Science*, 2(1995):209 –215.

Sotillo, M. G., Mourre, B., Mestres, M., Lorente, P., Aznar, R., Garcia-Leon, M., Liste, M., Santana, A., Espino, M., and Álvarez, E. (2021). Evaluation of the operational cmems and coastal downstream ocean forecasting services during the storm gloria (january 2020). *Frontiers in Marine Science*, 8:300.

Sperrevik, A. K. and Christensen, K. H. (2015). Constraining energetic slope currents through assimilation of high-frequency radar observations. *Ocean Science*, pages 237–249.

Stanev, E. V., Schulz-Stellenfleth, J., Staneva, J., Grayek, S., Grashorn, S., Behrens, A., Koch, W., and Pein, J. (2016). Ocean forecasting for the German Bight: From regional to coastal scales. *Ocean Science*, 12(5):1105–1136.

Stanev, E. V., Ziemer, F., Schulz-Stellenfleth, J., Seemann, J., Staneva, J., and Gurgel, K. W. (2015). Blending surface currents from HF radar observations and numerical modeling: Tidal hindcasts and forecasts. *Journal of Atmospheric and Oceanic Technology*, 32(2):256–281.

Steinhilper, E. and Gruijters, R. (2017). Border deaths in the Mediterranean: what we can learn from the latest data.

Stewart, R. H. and Joy, J. W. (1974). Instruments and Methods HF radio measurements of surface currents. *Current*, 21:1039–1049.

Stommel, H. (1989). The slocum mission. *Oceanography*, 2(1):22–25.

Storto, A., Falchetti, S., Oddo, P., Jiang, Y.-M., and Tesei, A. (2020). Assessing the impact of different ocean analysis schemes on oceanic and underwater acoustic predictions. *Journal of Geophysical Research: Oceans*, 125(7):e2019JC015636.

Storto, A. and Oddo, P. (2019). Optimal assimilation of daytime SST retrievals from SEVIRI in a regional ocean prediction system. *Remote Sensing*, 11(23):2776.

Tanajura, C. A., Mignac, D., de Santana, A. N., Costa, F. B., Lima, L. N., Belyaev, K. P., and Zhu, J. (2020). Observing system experiments over the Atlantic Ocean with the REMO ocean data assimilation system (RODAS) into HYCOM. *Ocean Dynamics*, 70(1):115–138.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research Atmospheres*, 106(D7):7183–7192.

Testor, P., Be, K., Mortier, L., Sea, T., and Basin, A. (2005). Modeling the deep eddy field in the southwestern Mediterranean: The life cycle of Sardinian eddies. *Geophysical Research Letters*, 32:7–10.

Testor, P., Meyers, G., Pattiaratchi, C., Bachmayer, R., Hayes, D., Pouliquen, S., Petit de la Villeon, L., Carval, T., Ganachaud, A., Gourdeau, L., et al. (2010). Gliders as a component of future observing systems. In *OceanObs'09*. OceanObs' 09.

Tintoré, J., Casas, B., Heslop, E., Vizoso, G., Pascual, A., Orfila, A., Ruiz, S., Renault, L., Juzà, M., Balaguer, P., et al. (2013). The impact of new multi-platform observing systems in science, technology development and response to society needs; from small to large scales.... In *International Conference on Computer Aided Systems Theory*, pages 341–348. Springer.

Tintoré, J., Lana, A., Marmain, J., Fernández, V., Casas, B., and Reyes, E. (2012). HF Radar Ibiza data from date 2012-06-01.

Tintoré, J., Pinardi, N., Álvarez-Fanjul, E., Aguiar, E., and Álvarez-Berastegui, Diego, e. a. (2019). Challenges for Sustained Observing and Forecasting Systems in the Mediterranean Sea. *Frontiers in Marine Science*, 6.

Tintoré, J., Lana, A., Marmain, J., Fernández, V.and Casas, B., and Reyes, E. (2020). HF Radar Ibiza data from date 2012-06-01 (Version 1.0) [Data set]. Balearic Islands Coastal Observing and Forecasting System, SOCIB.

Tintoré, J., Lana, A., Marmain, J., Fernández, V., Orfila, A., Observing, B. I. C., and Forecasting System, S. (2014). SOCIB EXP RADAR Sep2014 (Version 1.0) [Data set].

Torrado, H., Mourre, B., Raventos, N., Carreras, C., Tintoré, J., Pascual, M., and Macpherson, E. (2021). Impact of individual early life traits in larval dispersal: A multispecies approach using backtracking models. *Progress in Oceanography*, 192(May 2020).

Umlauf, L. and Burchard, H. (2003). A generic length-scale equation for geophysical turbulence models. *Journal of Marine Research*, 61(2):235–265.

Undén, P., Rontu, L., Järvinen, H., Lynch, P., Calvo, J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., Garcia-Moya, J. A., Jones, C., Lenderlink, G., McDonald, A., McGrath, R., Navascues, B., Nielsen, N. W., Ødegaard, V., Rodriguez, E., Rummukainen, M., Rõõm, R., Sattler, K., Sass, B. H., Savijärvi, H., Schreur, B. W., Sigg, R., The, H., and Tijm, A. (2002). High Resolution Limited Area Model, HIRLAM-5 Scientific Documentation. Technical Report December, SMHI, Norrkoping, SWEDEN.

Vandenbulcke, L., Beckers, J.-m., and Barth, A. (2017). Correction of inertial oscillations by assimilation of HF radar data in a model of the Ligurian Sea. *Ocean Dynamics*, pages 117–135.

Vargas-Yáñez, M., Juza, M., García-martínez, M. C., Moya, F., Balbín, R., Ballesteros, E., Muñoz, M., Tel, E., Pascual, J., Vélez-belchí, P., Salat, J., and Devlin, A. T. (2021). Long-Term Changes in the Water Mass Properties in the Balearic Channels Over the Period 1996 – 2019. *Frontiers in Marine Science*, 8(March):1–17.

Vignudelli, S., Birol, F., Benveniste, J., Fu, L. L., Picot, N., Raynal, M., and Roinard, H. (2019). *Satellite Altimetry Measurements of Sea Level in the Coastal Zone*, volume 40. Springer Netherlands.

Visbeck, M. (2018). Ocean science research is key for a sustainable future. *Nature Communications*, 9(1):1–4.

Wikle, C. K. and Berliner, L. M. (2007). A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1-2):1–16.

Wilkin, J. L. and Hunter, E. J. (2013). An assessment of the skill of real-time models of Mid-Atlantic Bight continental shelf circulation. *Journal of Geophysical Research: Oceans*, 118(April):2919–2933.

Xie, J., Counillon, F., Bertino, L., Tian-Kunze, X., and Kaleschke, L. (2016). Benefits of assimilating thin sea ice thickness from smos into the topaz system. *The Cryosphere*, 10(6):2745–2761.

Yan, Y., Barth, A., and Beckers, J. M. (2014). Comparison of different assimilation schemes in a sequential Kalman filter assimilation system. *Ocean Modelling*, 73:123–137.

Yu, P., Kurapov, A. L., Egbert, G. D., Allen, J. S., and Kosro, P. M. (2016). Variational assimilation of HF radar surface currents in a coastal ocean model off Oregon. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 129:394–400.

Zhang, W. G., Wilkin, J. L., and 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(3):119–133.

List of Figures

List of Tables