www.nautilos-h2020.eu

Organisations
University of Algarve (UAlg)
Hellenic Centre for Marine Research (HCMR)
Norwegian Institute for Water Research (NIVA)

D8.6

Model approach implementation and specifications

Date: September 30th. 2021

Doc. Version 2.0

10.5281/zenodo.7224676

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101000825 (NAUTILOS). This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Document Control Information

Settings	Value
Deliverable Title	Model approach implementation and specifications
Work Package Title	Data Management
Deliverable number	D8.6
Description	This deliverable sets the ground base activities developed to support the simulation experiments which will be developed in WP9. The OSSE concept is delineated, the characteristic for the Nature Runs and for the forecasting models are set and the assessment methods are defined.
Lead Beneficiary	University of Algarve
Lead Authors	Flávio Martins (UAlg)
Contributors	HCMR: George Triantafyllou, Kostas Tsiaras; NIVA: Trond Kristiansen, Elizaveta Protsenko; UAlg: Fernando Mendonça
Submitted by	August 31st. 2021
Doc. Version (Revision number)	1.0
Sensitivity (Security):	Public
Date:	30/09/2021
Doi	10.5281/zenodo.7224676

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Sabine Marty	Reviewer	Approve	17/09/2021
Antonio Novelino	Reviewer	Approve	28/09/2021
Gabriele Piere	Final Reviewer	Approve	30/09/2021

Revision	Date	Created by	Short Description of Changes
0.0	30/04/2021	Flávio Martins	Initial creation of the document
0.1	27/07/2021	Flávio Martins	“Skeleton” submitted
1.0	31/08/2021	Flávio Martins	First complete version submitted
2.0	30/09/2021	Flávio Martins	Final version submitted

Nature of the deliverable		
R	Report	X
DEC	Websites, patents, filing, etc.	
DEM	Demonstrator	
O	Other	

Dissemination level		
PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	

ACKNOWLEDGEMENT

This report forms part of the deliverables from the NAUTILOS project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000825. The Community is not responsible for any use that might be made of the content of this publication.

NAUTILOS - New Approach to Underwater Technologies for Innovative, Low-cost Ocean observation is an H2020 project funded under the Future of Seas and Oceans Flagship Initiative, coordinated by the National Research Council of Italy (CNR, Consiglio Nazionale delle Ricerche). It brings together a group of 21 entities from 11 European countries with multidisciplinary expertise ranging from ocean instrumentation development and integration, ocean sensing and sampling instrumentation, data processing, modelling and control, operational oceanography and biology and ecosystems and biogeochemistry such, water and climate change science, technological marine applications and research infrastructures.

NAUTILOS will fill-in marine observation and modelling gaps for chemical, biological and deep ocean physics variables through the development of a new generation of cost-effective sensors and samplers, the integration of the aforementioned technologies within observing platforms and their deployment in large-scale demonstrations in European seas. The fundamental aim of the project will be to complement and expand current European observation tools and services, to obtain a collection of data at a much higher spatial resolution, temporal regularity and length than currently available at the European scale, and to further enable and democratise the monitoring of the marine environment to both traditional and non-traditional data users.

NAUTILOS is one of two projects included in the EU´s efforts to support of the European Strategy for Plastics in a Circular Economy by supporting the demonstration of new and innovative technologies to measure the Essential Ocean Variables (EOV).

More information on the project can be found at: http://www.nautilos-project.eu.

COPYRIGHT

© NAUTILOS Consortium. Copies of this publication – also of extracts thereof – may only be made with reference to the publisher.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable is part of WP8 and set the basis for the development of the modelling activities in NAUTILOS. The actual modelling activities will be developed following this report, in WP9, and reported by D9.1, D9.2 and D9.3.

NAUTILOS aims at improving marine observations for physical, chemical, and biological essential variables through the development of a new generation of cost-effective sensors and samplers and its integration in observing platforms. This effort will contribute to democratise ocean observation and improve the human capacity of understanding, describing, and predicting the ocean behaviour. In NAUTILOS, the advances in the capability of simulating and predicting the ocean, produced by the new generation of sensors, will be assessed using the concept of Observing System Simulation Experiments (OSSE). The main conceptual idea behind an OSSE is the substitution of the reality, which is difficult and expensive to observe, by a numerical simulation of very good quality. This simulation is commonly known as the Nature Run (NR). The NR is used as if it was the reality, and observations are extracted from it. These are the so called "Synthetic Observations". In parallel, a forecasting model is setup, assimilating the synthetic observations. The "goodness" of the forecast is assessed against the NR, using a set of evaluation parameters and metrics. OSSE are thus "data denial" experiments.

In NAUTILOS, three simulation experiments using OSSE will be conducted: In the whole Mediterranean Sea, in the SW Iberian Coast and in the Hardanger fjord system in Norway. These systems are very different in their size, relevant physical phenomena, biogeochemical phenomena, and relevant timescales. Additionally, different models and assimilation methods will be used. Nevertheless, a common conceptual framework is proposed for the three cases, customizing implementation details for each region to accommodate the different natural phenomena and the differences in model setup. Two sets of experiments will be conducted: The Pre-NAUTILOS runs and the Post-NAUTILOS runs. Both sets of experiments assimilate synthetic observations from the NR. The type and characteristics of the synthetic observations used in each set of experiments will typify the observation panorama of that scenario. The Pre-NAUTILOS scenarios will characterize the present situation in terms of *in situ* and remote sensing observations. The Post-NAUTILOS scenarios will be characterized by a larger number of *in situ* observations, closer distributed in space, as well as in time. The precision of the observations is as important as their values for assimilation in ocean models. The Post-NAUTILOS scenarios will use foreseen precisions for the sensors being developed within the project. Also, the results from NAUTILOS are foreseen to improve calibration of remote sensing products, rendering more accurate measurements of remote sensing variables. This will also be accounted for when creating the Post-NAUTILOS scenarios.

The assessment of the experiment results against the NR is an essential part of the analysis. In this deliverable a set of metrics is defined to assess model results. These statistics and assessment parameters describe the nature of model results from both a spatial and a temporal point of view.

The methodology to extract the synthetic observations from the NR is also an important step in the preparation of the OSSE. In this deliverable general methodology is defined to extract the synthetic observations, which must be perturbed with uncertainties similar to those encountered in the real sensors.

The deliverable also includes preliminary results of the NR produced by the modelling systems of the three study sites. These results show the adequacy of the proposed modelling systems to proceed to the next phase of the simulation experiments.

List of Figures

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

Abbreviation	Definition
AC	Algerian Current
ADI	Alternating-direction implicit
AIS	Atlantic-Ionian Stream
AMC	Asia Minor Current
AVG	Average
AW	Atlantic Water
CC	Cilician Current
CMEMS	Copernicus Marine Environment Monitoring Service
CNR	Consiglio Nazionale delle Ricerche (National Research Council of Italy)
E	East
EMODNET	European Marine Observation and Data Network
EnKF	Ensemble Kalman Filter
EOFs	Empirical Orthogonal Functions
ERSEM	European Regional Seas Ecosystem Model
EU	European Union
FABM	Framework for Aquatic Biogeochemical Models
FM	Forecasting model

FR	Free run
FRS	Flow Relaxation Scheme
GEBCO	General Bathymetric Chart of the Oceans
GDP	Gross Domestic Product
H2020	Horizon 2020
HCMR	Hellenic Centre for Marine Research, Greece
IBM	Individual Based Model
IG	Ierapetra Eddy
INE	Instituto Nacional de Estatística (National Institute of Statistics, Portugal)
IS4DVAR	Incremental Strong constraint 4D-Variational
KF	Kalman Filter
L4	Satellite global analysed gridded data set
LEC	Libyo-Egyptian Current
LG	Lion Gyre
LIW	Levantine Intermediate Water
LPC	Liguro-Provencal current
LTE	Latakia Eddies
MDT	Mean Dynamic Topography
MMG	Mersa-Matruh Gyre
MMJ	Mid-Mediterranean Jet
N	North
NAO	North Atlantic Oscillation
NAUTILOS	New Approach to Underwater Technologies for Innovative, Low-cost Ocean observation
NR	Nature Run
NVE	Norwegian Water Resources and Energy Directorate
OCASO	Coastal Environmental Observatory of the Southwest of the Iberian Peninsula
OSSE	Observing System Simulation Experiment
OSTIA	Operational Sea Surface Temperature and Sea Ice Analysis system
PE	Pelops Eddy
POM	Princeton Ocean Model
RG	Rhodes Gyre
RMS	Root Mean Square

RMSD	Root Mean Square Deviation
RMSE	Root Mean Square Error
ROMS	Regional Ocean Modeling System
S	Salinity
SI	Super-Individuals
SAG	Southern Adriatic Gyre
SG	Gulf of Syrte Gyre
ShG	Shikmona Gyre
SLA	Sea Level Anomaly
SMS-Coastal	Simulation Management System for Coastal Hydrodynamic Models
SOMA	Algarve Operational Modelling and Monitoring System
SSH	Sea Surface Height
SST	Sea Surface Temperature
STD	Standard Deviation
SW	South-West
T	Temperature
TC	Thyrrhenian Current
UAlg	University of Algarve
W	West
WCG	Western Cyprian Eddy
WP	Work Package

1. INTRODUCTION

NAUTILOS aims at improving marine observations for physical, chemical, and biological essential variables through the development of a new generation of cost-effective sensors and samplers and its integration in observing platforms. This effort will improve the human capacity of understanding, describing, and predicting the ocean behaviour. A key step in that process is the use of numerical models to aggregate information from various sources and produce comprehensive descriptions and forecasts of the ocean. In NAUTILOS, the advances in the capability of simulating and predicting the ocean, produced by the new generation of sensors, will be assessed using the concept of Observing System Simulation Experiments (OSSE). Activities will be developed in two work packages. In work package 8, namely in task 8.3, the basis and the methodology for the experiments is defined. Also, part of this work package, is the identification and retrieval of base information needed to execute the experiments, as well as the creation of the so-called Nature Runs (NR) from which synthetic observations will be extracted. The results of model related activities in work package 8 are reported in this deliverable. Work package 9 follows directly, with the actual execution of the experiments. The experiments will be performed in three different geographic locations, using different models and different assimilation strategies. The Mediterranean Sea, the SW Iberian Coast and the Hardanger fjord system in Norway will be addressed. This variety of geophysical systems and modelling systems will make the analysis more robust and generic.

In this deliverable the geophysical and oceanographic characteristics of each geographic location are first described. The modelling systems that will be used in each location are then presented, including the forcing datasets that will be used in the experiments, the particular details on the operational processes and the operational cycles used to obtain the forecasts. The general concepts underlying the OSSE in NAUTILOS are then defined, including the methodologies to create the NR, and the metrics to analyse the results. The backbone of these definitions can be common to the three sites, although some specific characteristics need to be tailored to each system, due to the different processes, time, and space scales present in each site. Different sections of the methodology are thus dedicated to each site, after the common methodological section. Part of this deliverable is also the creation of the NR datasets for each system. The results obtained are presented for the three sites, including global data analysis and statistics. Subsequent data analysis and statistics will be developed in work package 9, as part of the simulation experiments.

2. Model and Site Description

This chapter is composed by three sections where the main geophysical and oceanographic features of the three study sites are described. This description is followed by a presentation of the modelling systems used in each system, the operational settings and the data sources used to drive the models.

2.1. POM-ERSEM in the Mediterranean

The Mediterranean is a semi-enclosed sea, connected through the Strait of Gibraltar to the Atlantic Ocean in the west and through the Dardanelles strait to the Sea of Marmara and the Black Sea in the northeast (see Figure 1). There are two distinct sub-basins (Western/Eastern) separated by the shallow Sicily Strait (~500 m) which limits water exchange, thus decoupling hydrodynamic and ecological conditions (Crise et al., 1999).

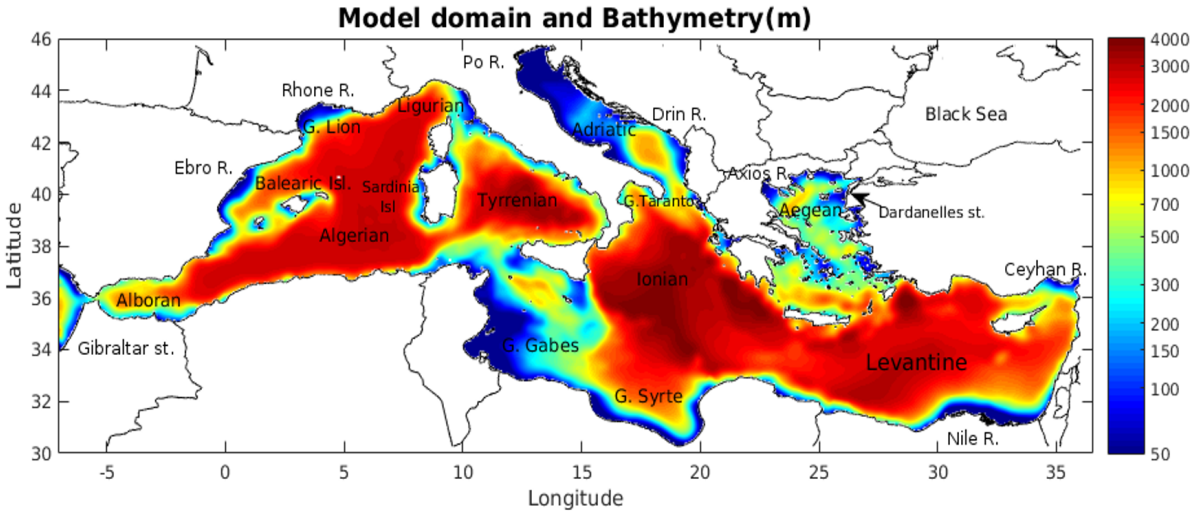

Figure 1 - Model domain and bathymetry (m, colour bar in logarithmic scale). Major rivers, straits and regional seas are indicated.

The net annual freshwater loss (Evaporation – Precipitation – Rivers – Black Sea inflow ≈ 3,250 km^3 yr^{-1}) is balanced by an inflow of Atlantic water at Gibraltar (Bryden et al., 1991). An anti-estuarine thermohaline circulation is driven by the negative freshwater and heat budget that transforms the relatively warm and fresh surface Atlantic Water (AW) to the colder and more saline Mediterranean intermediate water (Robinson et al., 2001). The main sites of dense water formation are the Gulf of Lions for the western basin and the southern Adriatic Sea, the north-eastern Levantine basin, and the Aegean Sea for the eastern basin (Laskaratos et al., 1999). The AW that is low in salinity and nutrients, follows an eastward pathway and reaches the central Levantine basin, where its density is increased during winter cooling and sinks (at ~300 m), forming the Levantine Intermediate Water (LIW), a water mass saltier and richer in dissolved nutrients (Theocharis et al., 1993). LIW follows a parallel course with AW, but moving in the opposite direction and eventually outflows into the Atlantic (Pinardi and Masetti, 2000).

In the surface, the western basin is characterized by an overall cyclonic circulation. The jet-like coastal Algerian current, surrounded by a series of smaller scale eddies, carries Atlantic water to the east, branching to the Atlantic-Ionian stream, through the Sicily channel and to the Thyrrenhian current to

the northwest (see Figure 2). The latter, followed by the Liguro-Provencal current, along the northwestern coast, forms a cyclonic circulation that is intensified in the G. Lions.

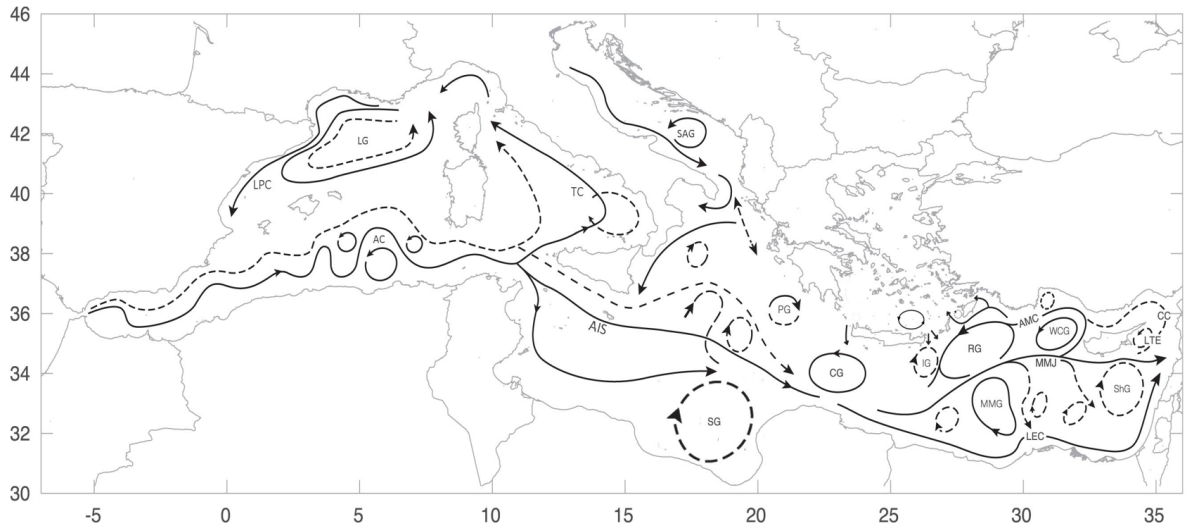

Figure 2 - Schematic general surface circulation of Mediterranean Sea. Dashed lines refer to recurrent or transient upper thermocline features. (LPC: Liguro-Provencal current; LG: Lion Gyre ; AC Algerian Current; TC: Thyrrhenian Current; SAG: Southern Adriatic Gyre; AIS: Atlantic-Ionian Stream; MMJ: Mid-Mediterranean Jet; CC: Cilician Current; AMC: Asia Minor Current; LEC: Libyo-Egyptian Current; SG: Gulf of Syrte Gyre; RG: Rhodes Gyre; MMG: Mersa-Matruh Gyre; ShG: Shikmona Gyre; IG: Ierapetra Eddy; PE: Pelops Eddy; WCG: Western Cyprian Eddy; LTE: Latakia Eddies).

The surface circulation in the Eastern Mediterranean is dominated by a series of semi-permanent jet-like currents and mesoscale eddies (Pinardi and Masetti, 2000; Robinson, 2001; Brankart and Brasseur, 1998). The Atlantic-Ionian Stream (AIS) crosses the Ionian Sea, bordered by an anticyclone to the southwest (Gulf of Syrte) and a cyclone to its northeast (Western Crete). AIS continues as the Mid-Mediterranean Jet (MMJ) that crosses the Levantine basin with a cyclonic pathway, either deflected to the west of Cyprus (during summer) or continuing to the Lebanese and Syrian coasts to create the Cilician (CC) and Asia Minor (AMC) currents along the Turkish coast to the north. Rhodes gyre is the most important cyclonic feature, while a series of semi-permanent anticyclonic (Mersa-Matruh, Shikmona) or re-current cyclonic/anti-cyclonic (Latakia) eddies are also observed, to the southeast area below the MMJ pathway. Other known features are also the Ierapetra, and Pelops anti-cyclones, while the Southern Aegean is occupied by the semi-permanent Cretan Sea cyclone and Western Cretan anti-cyclone (Theoharis et al., 1999). Recent observations from drifter campaigns (Menna et al., 2012) have also provide evidence for an along-slope cyclonic current that follows the southeast (Egyptian, Israel, Lebanon, Syrian) coastline to the north.

The Mediterranean pelagic ecosystem may be considered as oligotrophic, exhibiting a well-defined eastward decreasing gradient in plankton productivity (Moutin and Raimbault, 2002; Bosc et al., 2004; Kalaroni et al., 2020). This is related to the above described anti-estuarine circulation, with inflowing nutrient poor surface Atlantic water and outflowing subsurface Mediterranean waters. The primary production is mainly controlled by vertical mixing processes that supply the euphotic zone with deep water nutrients, reaching its maximum between December and April and minimum between June and September. The seasonal cycle is stronger in areas characterized by deep water formation, such as the Gulf of Lions in the northwestern Mediterranean, which is one of the most productive areas in the Mediterranean (Morel et al. 1991; Bosc et al. 2004). Relatively increased productivity is also found in areas receiving river nutrient inputs (see Figure 1), such as the Northern Adriatic, the North Aegean, and the Gulf of Lions.

A three-dimensional coupled hydrodynamic/biogeochemical model is implemented at Mediterranean basin scale (1/20° resolution ~5 km), building on the currently operational model within HCMR POSEIDON forecasting system (www.poseidon.hcmr.gr; Korres et al., 2007; Kalaroni et al., 2020). The hydrodynamic/biogeochemical model is also coupled with a Lagrangian particle drift model to track the fate of floating plastics from major source inputs.

The hydrodynamic model is based on the Princeton Ocean Model (POM), which is a three-dimensional, sigma-coordinate, free surface and primitive equation model. POM is a widely spread community model that has been used both for coastal and open ocean studies (e.g., Korres and Lascaratos, 2003; Kourafalou and Tsiaras, 2007; Zavatarelli and Pinardi, 2003) and has been extensively described in the literature (Oey et al., 1985; Blumberg and Mellor, 1987; Mellor, 2004). The model prognostic variables are temperature, salinity, velocity, sea surface height and turbulent kinetic energy. The model uses a bottom-following sigma coordinate system and an Arakawa-C staggered grid in the horizontal. The vertical eddy viscosity and diffusivity coefficients are computed using the Mellor-Yamada 2.5 turbulence closure scheme (Mellor and Yamada, 1982) that solves the equations for turbulent kinetic energy and turbulence macroscale, considering the wind stirring and the stability induced by stratification. Horizontal diffusion is calculated along sigma-levels following a Smagorinsky formulation (Smagorinsky, 1963). Time integration is performed with a split time step, in which the barotropic and baroclinic modes are integrated separately with a leapfrog time differencing scheme.

The hydrodynamic model employs a hybrid ensemble data assimilation scheme (Tsiaras et al., 2017) to correct the simulated near surface circulation, based on satellite altimetry (sea surface height) and sea surface temperature data. This is an ensemble based Kalman Filter (KF) that combines a flow-dependent error covariance, estimated from a stochastically generated ensemble (Hoteit et al., 2012), with a static background covariance, built from a set of Empirical Orthogonal Functions (EOFs), as described by Hoteit et al., (2001). Localization is applied as in Nerger et al. (2006), using only observations within a specified distance (radius ~50 km) from the updated grid point. The model state vector includes temperature, salinity, horizontal velocity and turbulence (kinetic energy, length scale), surface elevation and depth integrated horizontal velocities. Satellite data are assimilated every 8 days, assuming an observation error of ~0.03 cm for SSH and ~0.8 °C for SST, as in Korres et al. (2007). The ensemble size is N=101, including a flow-dependent ensemble with 8 members.

The biogeochemical model is based on the European Regional Seas Ecosystem Model (ERSEM; Baretta et al., 1995) that follows a "functional" group approach where the ecosystem is described in terms of functional roles (producers, consumers, decomposers). The pelagic plankton food web is adequately described with four phytoplankton groups (diatoms, nanoplankton, picoplankton, dinoflagellates), three zooplankton groups (heterotrophic nanoflagellates, microzooplankton, mesozooplankton) and bacteria. The pelagic model variables include also particulate and dissolved organic matter, along with dissolved inorganic nutrients (nitrate, ammonium, phosphate, silicate). Carbon dynamics are loosely coupled to the chemical dynamics of nitrogen, phosphate and silicate through a variable organic matter C:N:P:Si ratio scheme. The biogeochemical model is currently operational on basin scale, as part of the POSEIDON forecast system (Kalaroni et al., 2020).

The Lagrangian dispersion Individual Based Model (IBM) builds on Pollani et al. (2001) and takes into account the most important processes (advection from currents, stokes drift, vertical & horizontal mixing, biofouling/sinking, wind drag, beaching). The ocean currents and horizontal/vertical turbulent mixing coefficients are obtained on-line from the hydrodynamic model. The waves stoke's drift and mixing induced from waves are obtained off-line from a Mediterranean wave model output (see below). Different types and size classes of macro- (5-20 mm, 20-200 mm, >200 mm bottle/bag/foam) and microplastics (50 μm, 200 μm, 350 μm, 500 μm, 1000 μm, 2000 μm) are considered in the model. Biofouling induced sinking is explicitly described, as a possible mechanism of microplastics removal from the surface, due to the buoyancy loss resulting from the attachment of heavier biofilm. The wind drag that is practically effective only for macroplastics > 20 cm (bottles, foam) is assumed to depend

on the particle surface above water, following Yoon et al. (2010). The model follows the concept of Super-Individuals (SI) for computational efficiency. Each SI represents a group of particles, sharing the same attributes. The position of every SI is described by its coordinates (x, y, z) in a Cartesian system, which are updated every time-step using the 3-D displacement, produced by currents and wave, obtained with bi-linear interpolation at the SI location.

The 3-D hydrodynamic model domain covers the entire Mediterranean basin (7° W - 36° E, 30.25° N - 45.75° N) with a resolution of 1/20° × 1/20° (~5 x 5 km) in the horizontal and 24 sigma-levels in the vertical, following a logarithmic distribution approaching the surface and bottom layers. The General Bathymetric Chart of the Oceans (GEBCO; www.gebco.net, resolution: 0.4 x 0.4 km) was used to build the model bathymetry using bi-linear interpolation into the model grid. The hydrodynamic model is initialized with MODB-MED4 climatological sea temperature and salinity field data (http://modb.oce.ulg.ac.be/backup/modb/gridded.data.html; Brasseur et al., 1996), followed by a 5-year spin-up simulation.

The atmospheric forcing in the model is obtained from the POSEIDON operational weather forecast. The momentum, heat and freshwater fluxes at the air-sea interface were calculated using hourly fields of wind velocity (10 m), relative humidity (2 m), air temperature (2 m), precipitation, net incoming short-wave radiation and incoming long wave radiation, using a properly tuned bulk formulae set (Korres and Lascaratos, 2003).

The waves forcing (stokes drift, wave period and significant height) that is used in the Lagrangian drift model was obtained off-line from the Mediterranean waves hindcast (1/24°) that is available within Copernicus marine service (http://marine.copernicus.eu). The wave model is also a component of POSEIDON forecasting system and is based on the state-of-the-art third-generation wave model WAM Cycle 4.5.4 (Günther and Behrens, 2012).

The water exchange regime at Dardanelles straits was parameterized using a two-layer (Black Sea Water inflow) open boundary condition with prescribed transport rates and salinity (Nittis et al., 2006). The model has also an open boundary west of the Gibraltar strait. The implemented open boundary conditions are as follows: (a) zero-gradient condition for the free surface elevation; (b) a Flather (1976) boundary condition for the integrated (barotropic) velocity; (c) a Sommerfeld radiation condition for the baroclinic velocities; and (d) an upstream advection scheme for temperature, salinity and biogeochemical variables at the open boundary during outflow, while seasonal climatology is used to obtain temperature/salinity (MODB-MED4 data) during water inflow. Inorganic nutrients (phosphate, nitrate, silicate) are obtained from MEDATLAS climatology, while total alkalinity and dissolved inorganic carbon are obtained from GLODAP annual climatology.

The satellite data that were used for the data assimilation and validation of the Mediterranean simulated hydrodynamics were the Mean Dynamic Topography (MDT; spatial resolution: 0.0625° X 0.0625°) for the 2010-2018 period, obtained from the European AVISO+ altimetry data base (http://www.aviso.altimetry.fr; Rio et al., 2014) and daily Sea Level Anomalies (SLA; spatial resolution: 0.125° X 0.125°) that were obtained from the European Copernicus data base. Additionally, satellite remote sensing daily Sea Surface Temperature data (SST; spatial resolution: 0.04° X 0.04°) were used, obtained also from the Copernicus data base (Buongiorno et al., 2013; Pisano et al., 2016). To compare with the model outputs, all satellite data were interpolated on the model grid.

2.2. MOHID IN THE SW IBERIAN COAST

Located in the extreme west of the Iberian Peninsula, the District of Faro, commonly known as the Algarve, is the entire southern coastal strip of Portugal (Figure 3). Its territory covers almost 5,000 km², bordering the Region of Alentejo to the north and Spain to the east. It has a coastline at the south with approximately 160 km, which goes from Cape St. Vicente, at the westernmost tip, to the Spanish

border, and 50 km northward, at the west side, in the so-called Vicentina Coast, totaling a coast of around 200 km at the Atlantic Ocean. This region is divided into 16 municipalities, has a population of over 400,000 inhabitants and is one of the most important tourist destinations of Europe during summer (Nunes et al., 2020).

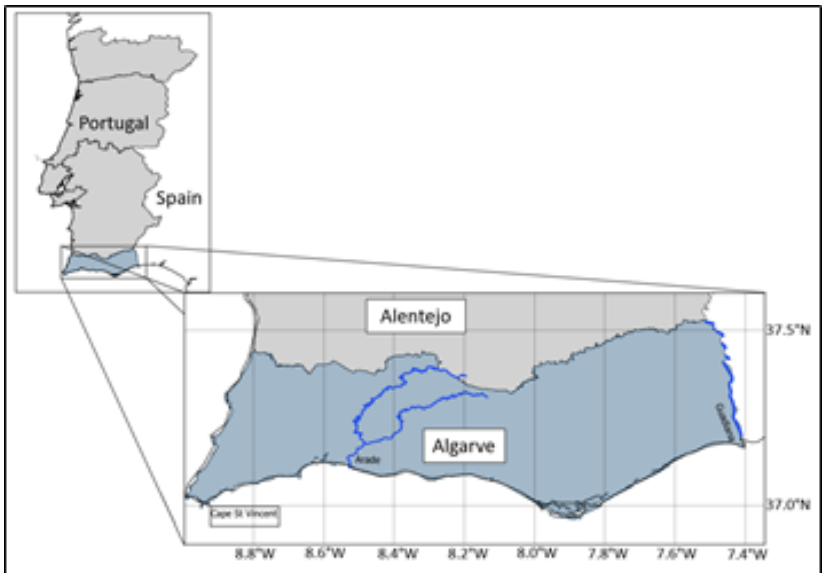

Figure 3- Geographic location of the Algarve.

The coast of the Algarve is formed mainly by rock-cliffs and sandy beaches, and it hosts three natural parks, which are of extreme importance for the region's biodiversity (Janeiro et al., 2012). The Ria Formosa Natural Park is one of the most important natural reserves of Portugal, with a total area of more than 18,000 ha. It is a wetland that hosts a wide diversity of habitats, serving as nursery ground for many marine species and a refuge for other numerous species of migrating birds (Aníbal et al., 2019). In addition to that, there is also the Natural Park of Southwest of the Algarve and Vicentina Coast, which is a natural reserve with unique biodiversity, that stretches for over 100 km in the coastline, over an area of 74,500 ha. Besides the Natural Parks, the Algarve coast also encompasses two important estuary areas of the Guadiana River and Arade River. They are subject to intense human activity, such as transport and trade, aquaculture and fisheries, salt production, and recreation (Moura et al., 2017).

The typical climate of the Algarve is Mediterranean, characterized by very hot and dry summers, and humid and mild winters. The average annual temperature is around 17 °C, but during the warmest months it usually presents values that exceeds 22 °C and may have maximums around 40 °C (Beck et al., 2018; Santo et al., 2005). In the colder months, the temperature is below 18 °C with minimums normally above 0 °C. In the summer, the region is often subject to moderate and extreme droughts, caused by the combination between the effects of the North Atlantic Oscillation (NAO) with the severe low precipitation rate (Dias et al., 2020). It is in the coastal zones where precipitation has the lowest rate, which decreases the average for the entire region, reaching less than 400 mm/year (Santos et al., 2010). The wind regime in Algarve is determined by the NAO: in the summer the difference between the north-south dipole of geopotentials anomalies is usually great, which favor westerlies winds in northern Europe. However, between March and August, the high-pressure field center also

dislocates towards Portugal, creating a displacement of northerly winds at the west coast (Fiúza et al., 1982). In the winter, the pressure gradient reduces producing weak westerly winds.

The coast of the Algarve has a semi-diurnal tidal cycle, resulting in two high and low water levels per day. The tidal range varies from 1.28 m in the lowest neap tide to a maximum of 3.44 m in the highest spring tide. The tide cycle is thus mesotidal, with a mean tidal range of 2.0 m of amplitude (IH, 2009). The wave climate varies from moderate to high (Costa et al., 2001). The mean wave height is between 1.7 to 2.2 meters with peak periods of nine seconds for summer and 13 for winter. Wave regimes are different for the south and west coast of the Algarve. In the south coast the dominant waves are westerly to southwesterly, while at the west coast, as it is exposed to the North Atlantic swell and storms, the dominant approach is from northwesterly to westerly with high energy condition and seasonality in wave climate (Costa et al., 2001).

The observed oceanographic patterns in the Iberian system reveal a distinct succession of mesoscale structures such as jets, meanders, ubiquitous eddies, upwelling filaments and countercurrents, superimposed on the more stable variations at seasonal timescales (Relvas et al., 2007). In the Algarve the upwelling season starts in the end of spring, stretches after the summer and is based on the cyclical wind regimes associated with the zonal displacement of the Azores high and Icelandic low-pressure systems (Fiúza, 1983; Haynes & Barton, 1990). At this event occurrence, strong northerly winds blow along the west coastline (Garel et al., 2016), pushing the coastal water away, which makes room for the resurgence of deep cold waters. The upwelling is also responsible by the transport of the surface waters offshore through an Ekman layer. In Portugal, the cold waterfront can reach from 30 to 50 km offshore in weak occurrences of upwelling and from 100 to 200 km, in the course of strong ones (Fiúza, 1983). The north winds can be diverted to the south coast of the Algarve due to the local topography, which also induces conditions for upwelling in this region. Under these conditions, the coastal flow along the southern Algarve moves eastward, as a result of west coast upwelling or of locally induced upwelling, while warmer oceanic waters lying offshore of the continental shelf flow westward.

Upwelling in the south coast of Algarve is more frequent during the hot season, with intensity and frequency decreasing from west to east along the coastline (Relvas & Barton, 2005). When meteorology conditions are not favorable for its formation, winds come predominantly from the west (Relvas et al., 2007). This also causes a change in the sea circulation, which makes possible the manifestation of a warm counter flow coming from the eastern Gulf of Cadiz to the southwest coast of the Iberian Peninsula (Teles-Machado et al., 2007). In the cold season, even though upwelling may appear, caused by the same northern winds, it is not common, and the water thermohaline properties are different from the summer (Álvarez et al., 2003). During the winter, as the direction of the dominant winds changes, so does the water flow, that assumes a poleward route carrying warm and saline water along the Portuguese coast (Frouin et al., 1990).

The Algarve Operational Modelling and Monitoring System, or simply SOMA, is the high-resolution operational model designed to reproduce the Portugal south coast dynamics. The model was developed based on the MOHID Modelling System environment (Leitão et al., 2005; Martins et al., 2001; Neves, 2007). MOHID is a robust numeric tool with modular architecture, programmed in ANSI FORTRAN and include features to simulate physical, chemical and biological processes of the marine ecosystem. This architecture configures the principle of object-oriented programming, which enables the simultaneous execution of several modules during a simulation and makes it a great tool for

downscaling applications (Braunschweig et al., 2004; Janeiro et al., 2014). Ocean dynamics in MOHID is simulated by solving the Navier-Stokes equations with the Eulerian approach, coupled with the finite volume method for space discretization. The discrete form of the equations is applied macroscopically to the control volume of each grid cell, then the model solves three-dimensional primitive equations for incompressible and compressible flows, in which hydrostatic equilibrium and Boussinesq approximation are assumed. The temporal discretization uses a semi-implicit alternating-direction implicit (ADI) algorithm with two levels of time per iteration (Martins, 1999; Martins et al., 1998). The modelling system also features modules to simulate Lagrangian transport and offers the great advantage of using generic vertical coordinates. This enables the use of any type of geometry in more than one subdomain in the vertical axis (Martins et al., 1998).

The SOMA system arose from the need to produce a high-resolution model in order to provide predictions of the sea state and the trajectory of oil spills on the Algarve coast (Janeiro et al., 2017). For that it used the MOHID oil spill model with the Lagrangian transport approach. Besides forecasting oil spills, it was also evaluated the efficiency of downscaling methods in determining the sources of oil leakage by combining backtracking simulation with vessel trajectories. As a result, SOMA is a validated model and is enabled to make predictions of the hydrodynamic behavior and water properties of the implemented region. In addition to traditional calibration and validation, it is also being continuously assessed dynamically under the Coastal Environmental Observatory of the Southwest of the Iberian Peninsula (OCASO) project (IP, 2006; Lorente et al., 2019).

The operational model encompasses two levels of increasing resolution, using bathymetric data from the European Marine Observation and Data Network (EMODnet). The first level uses a 2 km horizontal resolution grid and the second a 1 km grid. The implementation area is shown in Figure 4. Both have the same vertical spatial discretization with 50 layers of cartesian coordinates. At the open boundary, a Blumberg & Kantha (1985) condition is applied to the water level and a Flow Relaxation Scheme (FRS) (Martinsen & Engedahl, 1987) is used for velocity, salinity and temperature. The communication between the two levels is performed also using the FRS method, to relax the water level, the zonal and the meridional horizontal velocity components, through an eight-cell band adjacent to the lateral boundary. Simulation step time is 30 and 15 seconds for the first and second levels respectively.

At the open boundary, conditions for temperature, salinity, and current velocities are supplied by CMEMS-MERCATOR (GLOBAL_ANALYSIS_FORECAST_PHY_001_024) solution. Since this provider do not solve the tide explicitly, SOMA has an additional grid level above the previous ones. As it is called, the Level 0 is a simple 2D hydrodynamic model, forced by the FES2012 global tidal solution and has the single purpose of generating and supplying the tidal conditions to level 1. For the atmospheric forcing fields, SOMA uses the results of the regional weather forecast system SKIRON (Kallos et al., 1997; Papadopoulos et al., 2002), provided by Atmospheric Modeling and Weather Forecasting Group of The National and Kapodistrian University of Athens. It gives hourly data with a 5 km resolution of wind velocity components, air temperature, specific humidity, total cloud cover, sea level pressure, total precipitation, upward and downward long wave flux, evaporation, latent heat flux and sensible heat flux.

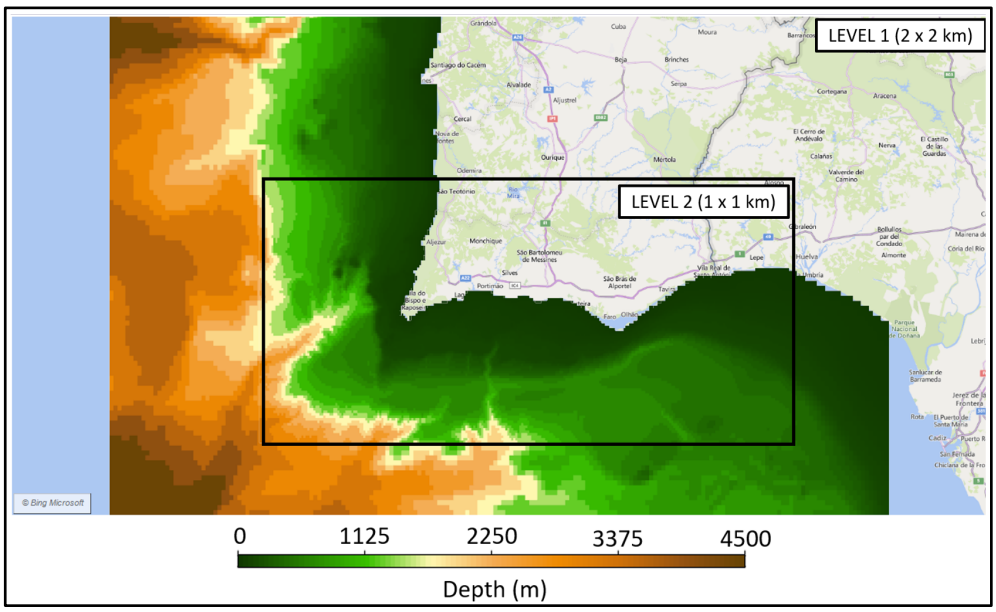

Figure 4 - SOMA bathymetry implemented according to EMODNET data.
Level 1 with a 2 km grid resolution and level 2 with 1 km grid.

SOMA is being operated in operational mode since July 2019 and is producing daily forecasts of the Algarve's coast hydrodynamics and water properties. Figure 5 shows examples of the system forecast simulation results. The model runs are being managed by the Simulation Management System for Coastal Hydrodynamic Models, or SMS-Coastal (Mendonça, 2020). The operationalization of SOMA consists in two types of simulation cycles: daily runs and weekly runs. The forecasts are obtained from the execution of the first type, which in addition to the external boundary conditions, also need the initial condition files generated in the previous day's cycle. To avoid excessive degradation along the sequence of daily runs, a restart is done once a week, based on assimilated initial conditions from CMEMS-MERCATOR (GLOBAL_ANALYSIS_FORECAST_PHY_001_024) solution, providing updated initial conditions. To prevent simulation instability, this weekly run is divided in two stages: the first is a two-day simulation with a time step of 10 and 5 seconds for Level 1 and 2; and the second is a four-day simulation with time step of 20 and 10 seconds. Variable relaxation is applied to gradually unconstrain the simulation. After this second stage, SOMA is fed with this solution to run its normal four-day forecast with the step time of 30 and 15 seconds for Level 1 and 2. The diagram shown in Figure 6 depicts the simulation cycles.

The model for the Algarve coast runs on a server computer, within Microsoft Windows 10 Enterprise 64-bit operating system, using 10 out of the 20 cores of an Intel Xeon Gold 6138 processor, 9.76 GB of RAM and an exclusive 450 GB data storage space. SOMA is being executed coupled with SMS-Coastal since July 7th, 2019, and after two years the system has managed 125 weekly runs and 770 daily runs, totaling 895 executions of the simulation cycles. From those executions a total of 80 failed runs were identified, as indicated by the program's log files, which corresponds to less than 9 % of the total runs. Even so, not all the failures prevented the model of running. As indicated in the "Restarted run" column of Table 1, in two of the cases the model was restarted after the failure was identified. Also, in the table, the failures were classified by each simulation cycle and by categories, which are:

- Code error: SMS-Coastal crash caused by poor programming of a new module or in a code update.

- Manual emergency stops: execution manually aborted by the user.
- Run terminated by MOHID: SMS-Coastal recognizes that the MOHID executable unsuccessfully ended, due to model instability, lack of external forces, or anything else presented in MOHID execution log.
- Restart files not found: SMS-Coastal was unable to prompt a daily run cycle due to the lack of restart files.
- Computer problems: SMS-Coastal and/or MOHID stopped due to insufficient RAM, execution window frozen or unscheduled maintenance reboot.

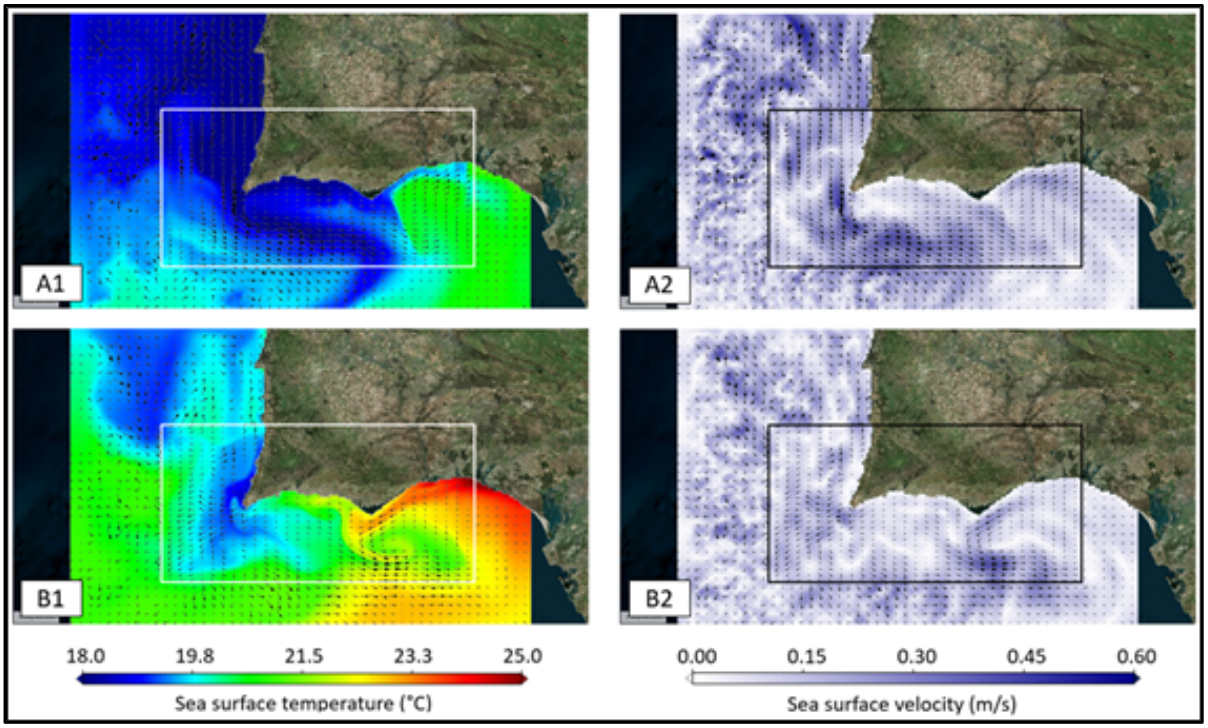

Figure 5- Results of SOMA forecast simulations. In "A" an insight from June 19th 2020, at 12:00 pm, of the sea surface temperature (A1) and velocity (A2). The upwelling is very well defined and it also extends to the south coast. Almost one month later (B) in July 18th, 12:00 pm, the upwelling is losing its strength and giving room to the counter current from the Gulf of Cadiz.

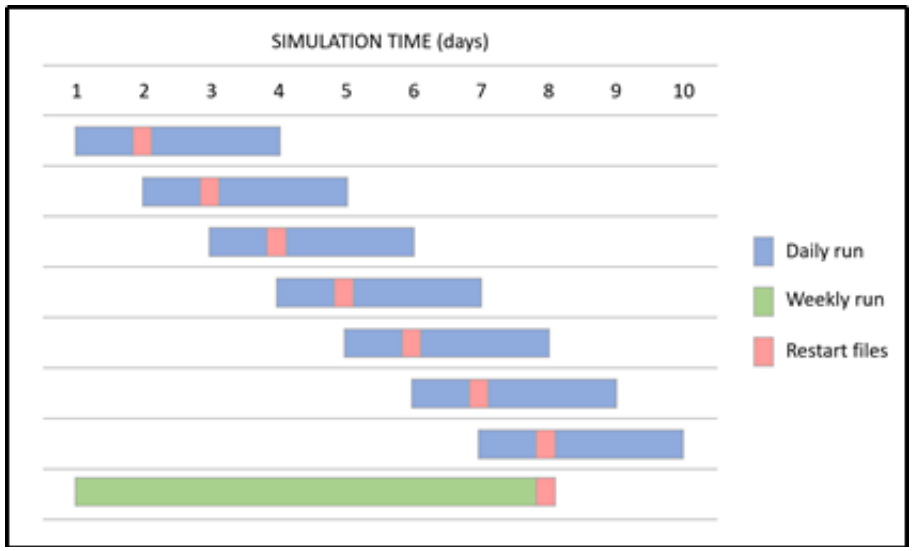

Figure 6 - Operational cycle assuming a four-day forecast: a new daily run starts (blue ribbon) and, after a day of simulated time, the model results are used as initial conditions for the next day simulation (red ribbon). After seven daily cycles, a weekly run (green ribbon) is executed in hindcast mode to provide the initial conditions for the simulation of the eighth day. In that day the restart files from the weekly run take priority over the daily simulation files that run in parallel.

Table 1- Failures count and proportion from July 7th, 2019, to July 7th, 2021.

Failures	Restarted run	Weekly run	Daily run	Total	% in failures	% in total runs
Code errors	Yes	3	12	15	18,8%	1,68%
Manual emergency stops	Yes	2	7	9	11,3%	1,01%
Run terminated by MOHID	No	12	24	36	45,0%	4,02%
Restart files not found	No	0	7	7	8,8%	0,78%
Computer problems	No	4	9	13	16,3%	1,45%
Totals		**21**	**59**	**80**	**100%**	**8,94%**

2.3. ROMS-ERSEM IN THE HARDANGER FJORD SYSTEM

The Hardangerfjord is the second longest fjord in Norway stretching 179km from the Atlantic Ocean to its innermost section. It is also one of the most productive fjords in Norway for salmon farming with an annual production of 70K tonnes (Husa et al. 2014). Over the last decades the Hardangerfjord has been influenced by a variety of anthropogenic stressors such as hydroelectric powerplants, industry, and fish farming. To ensure good ecological status of the fjord, a national assessment program continues to monitor the physics, chemistry, nutrients and biology of the fjord system, which last year was assessed as “good environmental status”. Still, with increasing ocean temperatures and continued high or increased fish farming it is vital for the overall health of the fjord to be closely monitored.

The ROMS-ERSEM model is based on an existing modelling system (Palmer et al. 2019) developed by the Norwegian Institute for Water Research. The modelling system has previously been used in other projects, but as part of NAUTILOS will be upgraded with increased spatial resolution, improved spatial resolution of the atmospheric and ocean lateral forcing, and implementing assimilation of satellite

data. The model system couples the ERSEM biogeochemical model (Butenschön et al., 2016) to describe the ecosystem dynamics with the Regional Ocean Modeling System (ROMS) physical model using the Framework for Aquatic Biogeochemical Models (FABM, Bruggeman and Bolding, 2014). The model configuration is implemented as a fjord-scale model at 160 m resolution covering the Hardangerfjord region (hereafter "ROHO160", Figure 7).

For the ROHO160 hindcasts, initial and boundary conditions for the ocean physics (temperature, salinity, circulation) were provided by GLORYS12V1 (Lellouche et al. 2018). GLORYS12V1 is a global, higher resolution model that assimilates available observations from buoys, ships and satellites to produce a historical reanalysis covering the temporal period 1993-2021, with a spatial resolution of 1/12° degrees. To produce the required forcing file for ROMS based on GLORYS12V1 we applied the Python toolbox model2roms (DOI:10.5281/zenodo.5140153).

Tidal forcing is based on the global ocean tidal model TPXO7.2 from Oregon State University, by imposing surface elevation and corresponding barotropic velocity components at the open boundaries. Freshwater input from rivers is from monthly averaged values of river-runoff from the Norwegian Water Resources and Energy Directorate (NVE).

The AROME global high-resolution atmospheric re-analysis archive has been applied as atmospheric forcing, providing information on atmospheric variables at 6-hour temporal resolution (00, 06, 12 and 18 UTC). AROME is provided by the Norwegian Meteorological Institute (www.met.no). Variables include pressure, humidity, cloud cover, surface winds, temperature, and precipitation. Longwave and shortwave radiation terms are analytically calculated internally using the bulk flux formula.

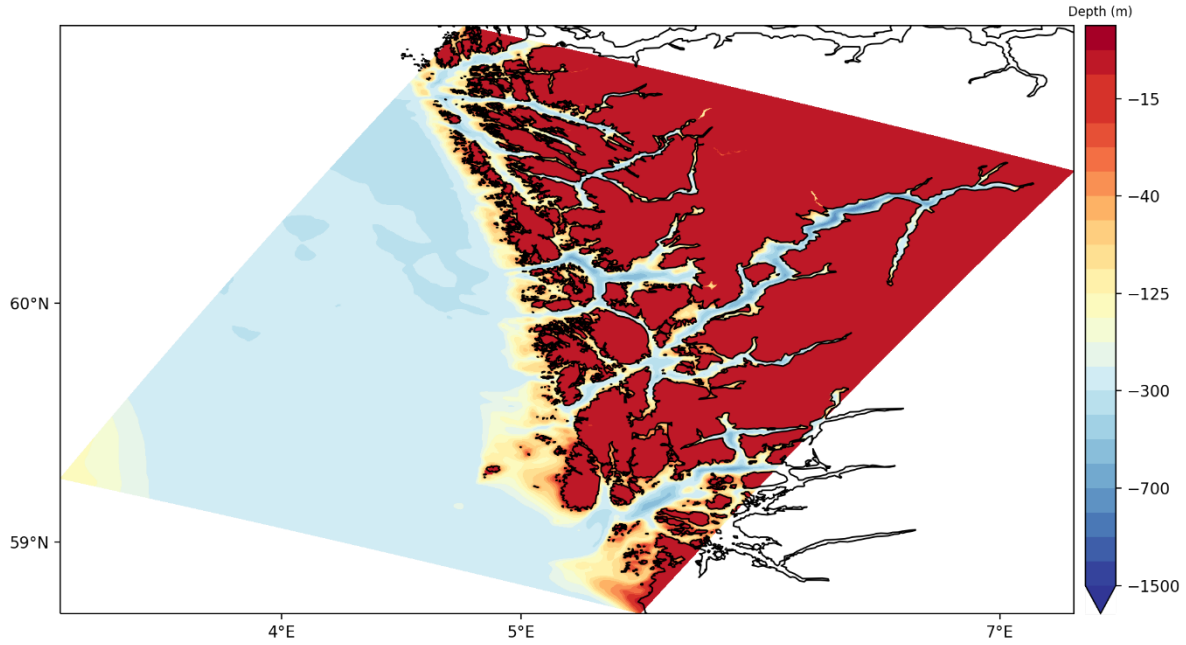

Figure 7 - Bathymetry of the Hardangerfjord model at 160 m resolution shows the complexity of the very deep fjord system surrounded by a myriad of small islands and land.

ROHO160 is run on a Norwegian national supercomputer (Sigma2) and requires extensive CPU usage and memory consumption. We currently run the model on 256 CPU and the speed of the simulations depends on whether biology is coupled or not. The ERSEM model has not been parallelized and can increase the time required for simulations by 7-fold. The simulations are performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway (#nn9297k).

3. OSSE METHODOLOGY IN NAUTILOS

In this chapter, the methodology used to implement the Observing System Simulation Experiments (OSSE) within the NAUTILOS framework is defined. The common approach and definitions are described in the first section, where the problem is put in perspective, the general methods to create the Nature Run are advanced, and the general metrics to assess the results are proposed. The first section is followed by three sections, one for each study site, where the details for each geographic location are specified. This includes the duration of the experiments and the details on their execution.

3.1. COMMON APPROACH

3.1.1. General OSSE Concept

One of the main purposes of ocean observations is to force and condition ocean models. It would thus be relevant to assess the impact of new observations on model forecasting capabilities, prior to the deployment of new observation missions. This assessment is a difficult task since the observations do not exist yet, and it is very expensive to set up ocean trials for that purpose. Acknowledging that, the concept of Observing System Simulation Experiments (OSSE) has been explored, starting in the decade of 1980's in the numerical weather prediction community (e.g., Atlas et al, 1985 and Atlas, 1997) and transiting progressively to the ocean forecasting community (e.g., Mourre et al., 2006 and Oke and Schiller, 2007). The main conceptual idea behind an OSSE is the substitution of the reality, which is difficult and expensive to observe, by a numerical simulation of very good quality. This simulation is commonly known as the Nature Run (NR). The NR is used as if it was the reality, and observations are extracted from it. These are the so called "Synthetic Observations".

In parallel, a forecasting model is setup, assimilating the synthetic observations. The "goodness" of the forecast is assessed against the NR, using a set of evaluation parameters and metrics. OSSE are thus "data denial" experiments. As highlighted by Halliwell et al. (2014, 2015) the numerical characteristics of the NR and the Forecasting Run should be as different as possible, including, when possible, different models, with different "numerics", different resolutions, different grids, different parameterizations, and different forcing. The objective is to resemble in these numerical experiments, as much as possible, what is done during real observation, where we are observing the nature which is essentially different from any model. Naturally, it is difficult to guarantee all this diversity between NR and Forecasting model. Many OSSE are thus based in the same model "numerics", changing as much as possible the other factors. This type of experiments is commonly termed "fraternal twin experiments". Although different, both the NR and the forecasting model results should share similar "statistics" with the real observations, to assure they represent correctly the physical phenomena being studied (Halliwell et al., 2014).

3.1.2. Common OSSE setup in NAUTILOS

As already anticipated, In NAUTILOS three simulation experiments will be conducted using OSSE: in the whole Mediterranean Sea, in the SW Iberian Coast and in the Hardanger fjord system in Norway. These systems are very different in their size, relevant physical phenomena, biogeochemical phenomena, and relevant timescales. Different models and assimilation methods will also be used. A common methodology is difficult to accommodate with such variety. Even though, in this section a common conceptual framework is proposed for the three cases. The implementation details are left open to accommodate the different nature of each region and the differences in model setup. The

subsections following this, will then detail the particular elements of each implementation. Figure 8 summarizes the common framework of the OSSE that will be developed in WP9. The concept of "fraternal twin" experiments will be used, meaning that the same numerical method will be used both for the NR and for the forecasting model. The needed diversity will be obtained from different model setups. The NR will have a spatial resolution which will be higher than the resolution used in the forecasting run. Additionally, the forcing and the parameterizations used can be different. The NR is run from t = 0 to t = T_f, either assimilating real observations or restarting periodically from global circulation models which assimilate real observations. This is still common practice for local and regional models.

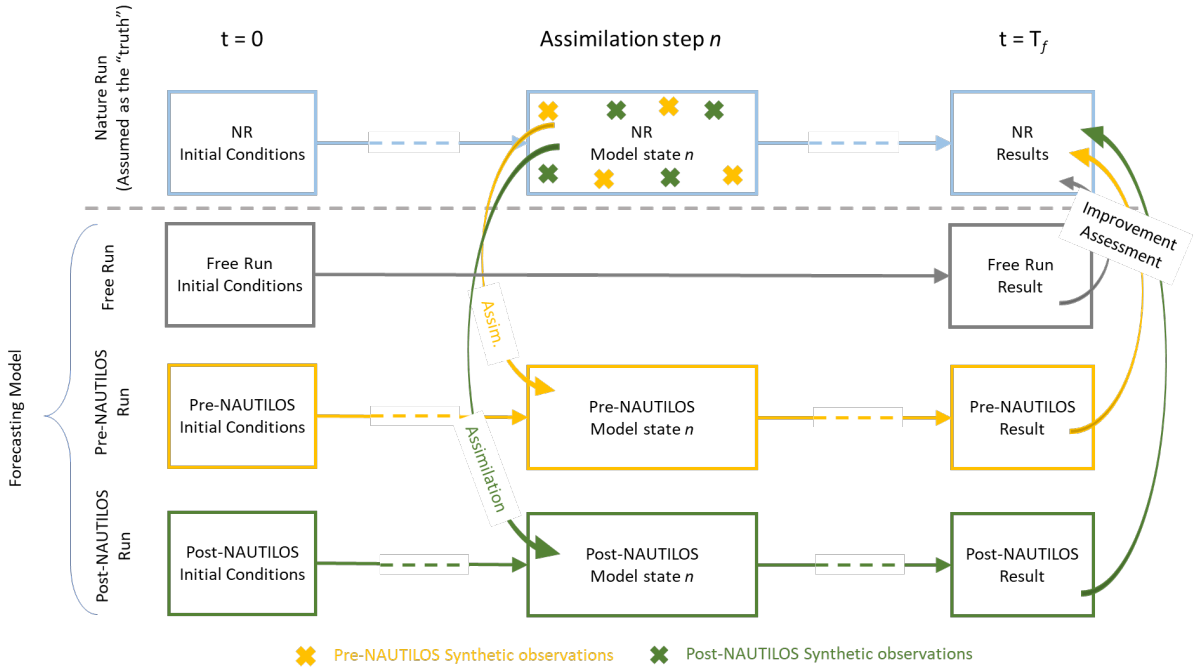

Figure 8 - Common framework used in NAUTILOS for the implementation of the Observing System Simulation Experiments (OSSE).

The duration of the experiments will be different for each site as the dimension of the systems are different and the timescales of the processes are also different.

A free run of the forecasting model, without any assimilation is then executed. This is the reference forecasting run against which improvements due to assimilation of synthetic observations are compared. The analysis of the free run is also important to assure that NR and forecasting model, though using different assumptions and producing different results, share the same common "statistics" of the natural phenomena under consideration.

Two sets of experiments will then be conducted: The Pre-NAUTILOS runs and the Post-NAUTILOS runs. Both sets of experiments assimilate synthetic observations from the NR. The type and characteristics of the synthetic observations used in each set of experiments will typify the observation panorama of that scenario. The Pre-NAUTILOS scenarios will characterize the present situation in terms of *in situ* and remote sensing observations. The essence of NAUTILOS is the democratization of ocean observations, thus the Post-NAUTILOS scenarios will be characterized by a larger number of *in situ* observations, closer distributed in space, as well as in time. The precision of the observations is as important as their values for assimilation in ocean models. The Post-NAUTILOS scenarios will use foreseen precisions for the sensors being developed within the project. Also, the results from NAUTILOS are foreseen to improve calibration of remote sensing products, rendering more accurate

measurements of remote sensing variables. This will also be accounted for when creating the Post-NAUTILOS scenarios.

As referred above, the processes and timescales are different for each site, and thus the particular details for the planning of each simulation experiment are specified in the sections below. Nevertheless, indicative properties and experiment characteristics are outlined in Table 2.

Table 2 - Indicative simulation experiment characteristics.
Other values and details will be defined for each site.

Simulation Characteristic	**Possible Values**	**Notes**
Time length	Annual, Seasonal, or Monthly	Depends on the size and processes of the system
Assimilated SST (Subskin) (Pre-Nautilos)	spatial resolution: 0.02° x 0.02° Temporal resolution: Daily Refresh: 18:00 Mean Bias: -0.03 K Std. Dev.: 0.45 K	Based on the CMEMS product: SST_EUR_PHY_L4_NRT_010_031 https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=SST_EUR_PHY_L4_NRT_010_031
Assimilated SST (Subskin) (Post-Nautilos)	spatial resolution: 0.02° x 0.02° Temporal resolution: Daily Refresh: 18:00 Mean Bias: *to be defined* Std. Dev.: *to be defined*	Same as the Pre-Nautilos SST since the satellite missions will be the same, only calibration is foreseen to be improved.
In Situ T and S profiles (Pre-Nautilos)	Positions: Present buoys available in each system producing operational sustained data Temporal resolution: Near Real Time Mean Bias: *to be defined* Std. Dev.: *to be defined*	Same mean bias and Std. Dev. For all systems.
In Situ T and S profiles (Post-Nautilos)	Positions: Present buoys Plus up to 300% increase due to Nautilos Technologies with increased depth range. Temporal resolution: Near Real Time Mean Bias: *to be defined* Std. Dev.: *to be defined*	

3.1.3. Common assessment methodology

As indicated in Figure 8 the results from the experiments using the forecasting model must be compared with the NR. These comparisons will enable to assess the expected improvements to be

obtained from NAUTILOS. Many indicators and metrics have been proposed for that purpose (e.g., Murphy, 1988, Taylor, 2001, Liu and Weisberg, 2011), along with more traditional assessment methodologies (e.g. RMSE analysis). Since the three sites possess different characteristics, the appropriate metrics can vary from site to site. Additionally, the appropriateness of the metrics can only be evaluated during the assessment phase, after the experiments are conducted. Below, a list of possible metrics is proposed, to be selected or complemented for each site according to the results.

Spatially Integrated Averages and Standard Deviations

Instantaneous spatial averages (AVG) and standard deviations (STD) of both the NR and the forecasting model results can be used to plot timeseries of model evolution:

$$AVG(t_i) = \frac{\sum_{domain} \theta(t_i)}{N} \quad (1)$$

$$STD(t_i) = \sqrt{\frac{\sum_{domain} \left(\theta(t_i) - AVG(t_i)\right)^2}{N}} \quad (2)$$

Where θ is the variable under analysis, from the forecast model or the NR in time t_i, and N is the number of cells used. θ can be instantaneous model results or averaged results over a time bin centred in t_i (e.g., daily averages). The length of the bin depends on the process and the site being analysed. The periodicity of the sampling can be equal or larger than the time bin (e.g., daily, weekly) and also depends on the process and the site being analysed. t_i spans the entire length of the experiment, and the spatial integration (sum) can be performed over the entire 3D field or over any 2D layer (e.g., the surface).

Time series obtained from these two metrics are considered useful to compare NR with forecast model behaviour, in a global way, along time. The same analysis can be performed using observed data (e.g. using L4 SST or SSH data products) to evaluate model statistics against nature. The objective here is not to calibrate the models but to ensure the models present statistics which are similar to those found in nature.

Spatially Integrated RMSD

Instantaneous spatially integrated root mean square deviation (RMSD) between the NR and the forecasting model results can be used to plot timeseries with evolution of model comparisons:

$$RMSD(t_i) = \sqrt{\frac{\sum_{domain} \left(\theta_f(t_i) - \theta_{NR}(t_i)\right)^2}{N}}$$

Where θ_f and θ_{NR} are the forecast and NR variable being assessed in time t_i and N is the number of cells used. They can be instantaneous model results or averaged results over a time bin centred in t_i (e.g., daily averages). t_i spans the entire length of the experiment and the sampling periodicity depends on the process and the site being analysed. (e.g., daily, weekly). Spatial integration (sum) can be performed over the entire 3D field or over any 2D layer (e.g., the surface). If the NR resolution is different from the forecast resolution, an average or any other type of mapping of the higher resolution grid into the lower resolution grid must be performed prior to the analysis.

The time series of this quantity is seen as a good first indicator of the forecast evolution in relation to the NR.

Temporally Integrated Averages and Standard deviations

The indicators referred in the previous section allow a global overview of the temporal evolution of the solutions, but do not allow a spatial evaluation. In the next sections, temporally integrated

indicators for each grid point allow a spatial evaluation of model behaviour. The temporally integrated average can be computed by:

$$AVG(i,j,k) = \frac{\sum_{t_i=0}^{T} \theta(i,j,k)}{N} \quad (3)$$

Where θ is the variable under analysis, from the forecast model or the NR in grid point i, j, k and T is the number of time instants considered. θ are either instantaneous model results or averaged results over a time bin centred in t_i (e.g., daily averages). The same logic can be used to compute the standard deviations:

$$STD(i,j,k) = \sqrt{\frac{\sum_{t_i=0}^{T} \left(\theta(i,j,k) - AVG(i,j,k)\right)^2}{T}} \quad (4)$$

From the two indicators above, horizontal maps per layer or vertical cuts can be plotted, providing a general idea of the spatial behaviour of model statistics.

Temporally Integrated RMSD

Using the same rationale, the root mean square deviation between the forecast and the NR can be computed:

$$RMSD(i,j,k) = \sqrt{\frac{\sum_{t_i=0}^{T} \left(\theta_f(i,j,k) - \theta_{NR}(i,j,k)\right)^2}{T}} \quad (5)$$

Allowing in the same way, the plotting of horizontal maps per layer or vertical cuts, which give a general idea of the spatial distribution of forecast deviations relative to the NR.

Murphy Skill Score

The abovementioned indicators provide a general overview of NR and forecast model behaviour, as well as a simple mean of comparing the two. Other more powerful statistics can also be used, as suggested in the literature. Maps of forecast skill can be constructed using the Murphy (1988) skill score:

$$\Sigma = \rho^2 - \left[\rho - \frac{\sigma_f}{\sigma_{NR}}\right]^2 - \left[\frac{(\bar{\theta}_f - \bar{\theta}_{NR})}{\sigma_{NR}}\right]^2 \quad (6)$$

Σ is applied to every grid cell during the analysed time period, to produce the skill map. ρ is the correlation between NR and the forecast, σ is the standard deviation and $\bar{\theta}$ is the average of the property being analysed. Note that Σ can assume negative values. According to Murphy (1988), values of $\Sigma < 0$ denote a "not significant" skill while values of $\Sigma > 0$ correspond to "significant skill". A "perfect skill" would correspond to a value of $\Sigma = 1$.

Taylor Diagrams

Taylor diagrams (Taylor, 2001) can be used to combine different statistics regarding the comparison of the forecast and the NR in a single diagram. Figure 9 represents a sample Taylor diagram. Forecast model results are represented by the coloured circles and triangles and are compared with the NR result, represented by the black circumference. The radial distance to the centre measures the standard deviation of model results, normalised by the standard deviation of the NR. The angle of the

dotted isolines above the horizontal measures the correlation between the model forecast and the NR and the grey isolines centred in the NR black circle measure the RMS difference of the forecast relative to the NR, normalized also by the standard deviation of the NR.

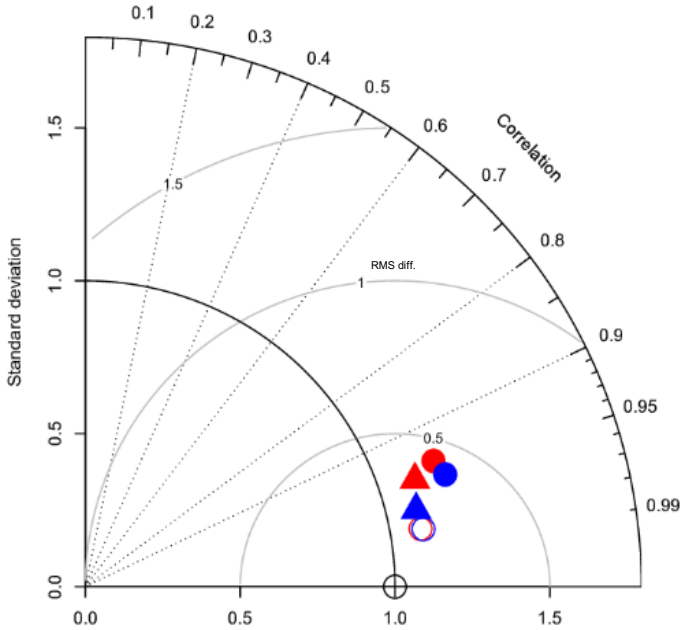

Figure 9 - Sample of a simple Taylor diagram. Forecast model results represented by the coloured circles and triangles are compared with the NR result, represented by the black circumference.

This diagram can be applied to spatially integrated time series of model variables (e.g., T, S, u, v). Generically, the closer the forecast result is to the NR the better is the model performance. The three statistics can however be evaluated separately, giving a broader insight of the model behaviour.

3.1.4. Common methodology to create synthetic observations

In the framework of OSSE, NR results will be used to extract synthetic observations which will then be used in the forecasting models. The objective is to emulate real observations using the NR as an "hypothetical truth". Like real observations, synthetic observations must also contain observation errors. These observations errors must be added to the NR values after extraction and must possess error levels and behaviour similar to those encountered in equivalent real observations using real instruments. Additionally, for some types of observations, representation or mapping errors should also be present. As pointed by Halliwell et al. (2014), when observations have correlation scales larger than the model resolution, representation errors must also be accounted. An example of this is for the along-track satellite observations resolving frontal processes or mesotidal eddies close to the cut off limit of model resolution. In those cases, along-track correlated errors must be added in a progressive along-track segment. For the purpose of NAUTILOS these are very specific situations lying outside the scope of NAUTILOS observations and sensors and will thus be discarded in our approach.

The sensors (instruments) observation errors will thus be considered uncorrelated. Also, errors will be assumed as gaussian. A gaussian random generator will be used to include a specified variability upon extraction of data from the NR. The standard deviation of that perturbation must be consistent with that of existing sensors or of new sensors being developed in NAUTILOS. Bias may also be included in this perturbation if applicable.

3.2. APPLICATION IN THE MEDITERRANEAN

The general methodology described in section 3.1 will be applied at basin-scale in the Mediterranean Sea, using the above-described hydrodynamic model (POM) that is based on the HCMR POSEIDON operational system. The upgraded hydrodynamic model (POM20) has a higher horizontal resolution (1/20^o ~5 km), as compared to the operational POSEIDON model (POM10) (1/10^o~10 km, Korres et al., 2007) and employs a hybrid (Hybrid) ensemble data assimilation scheme (Tsiaras et al., 2017) to correct the simulated near surface circulation, based on satellite altimetry (sea surface height) and sea surface temperature data. The higher resolution model (POM20) will provide the "nature run" (NR) dataset that will be used to extract synthetic observations, which will be assimilated by the forecasting model (FM) in the OSSE experiments (see above section 3.1). The forecasting model will be based on the lower resolution POM10 and will have a slightly different parameterization (e.g. horizontal/vertical mixing parameters) to ensure that NR and FM simulations are sufficiently different, resembling the error statistics between an ocean model and the true ocean, otherwise the OSSE experiments may lead to biased assessments, typically an overestimation of impact when sparse data are assimilated (Halliwell et al., 2014, see above section 3.1).

The Hybrid assimilation scheme (Tsiaras et al., 2017) will be used to assimilate the synthetic observations in the OSSE experiments. This is an ensemble based Kalman Filter (KF) that combines a flow-dependent error covariance with a static background covariance (see above section 2.1). The filter background static covariance, which represents a climatology of the system statistics, is obtained performing an empirical orthogonal functions (EOFs, Lorenz 1956; Hoteit et al., 2001) analysis on a long sequence of the FM model (POM10) output (e.g. 2 years). The initial ensemble was then sampled along the directions of the resulting EOFs (Hoteit et al., 2001). The filter operates as a succession of three steps: a "sampling" step that generates an ensemble of members describing the ocean state from an initial estimate and its covariance, a "forecast" step that propagates the generated ensemble members forward with the model and the "analysis" step, where the forecast mean and its covariance are updated, using the available observations. This assimilation "cycle", updating the forecast state of the model, will be repeated on a weekly basis (using all available or synthetic observations during the current week), as currently performed by the operational POSEIDON model. An example of available observations, assimilated by the operational POSEIDON model is shown in Figure 10 for January 2016. These are obtained from the Copernicus INSITU TAC database (http://www.marineinsitu.eu) and mainly consist of temperature and salinity (T, S) profiles, obtained from Medargo floats and Gliders.

In the OSSE experiments, synthetic observations will consist of (T, S) vertical profiles and potentially also sea surface temperature, extracted from the NR. These pseudo-observations will mimic existing or future in situ measurements from various platforms (e.g. drifters, profilers) and will be perturbed using a random function with a normal distribution, and an accuracy (observation error) equivalent to that of the real sensors. A series of Pre-Nautilos and Post-Nautilos OSSE experiments will be performed, adopting different accuracies and spatial coverage of the synthetic observations, based on the new technologies developed in Nautilos, foreseeing an increase in the number of buoys and sensors deployed. The OSSE simulations will be performed over a one-year period to cover the system seasonal variability, with particular focus on winter vertical mixing events that are particular relevant with plankton productivity (to be examined in T9.2). The forecast model in the OSSE will be initialized from a different year, as compared to the nature run or from a mean multi-year average.

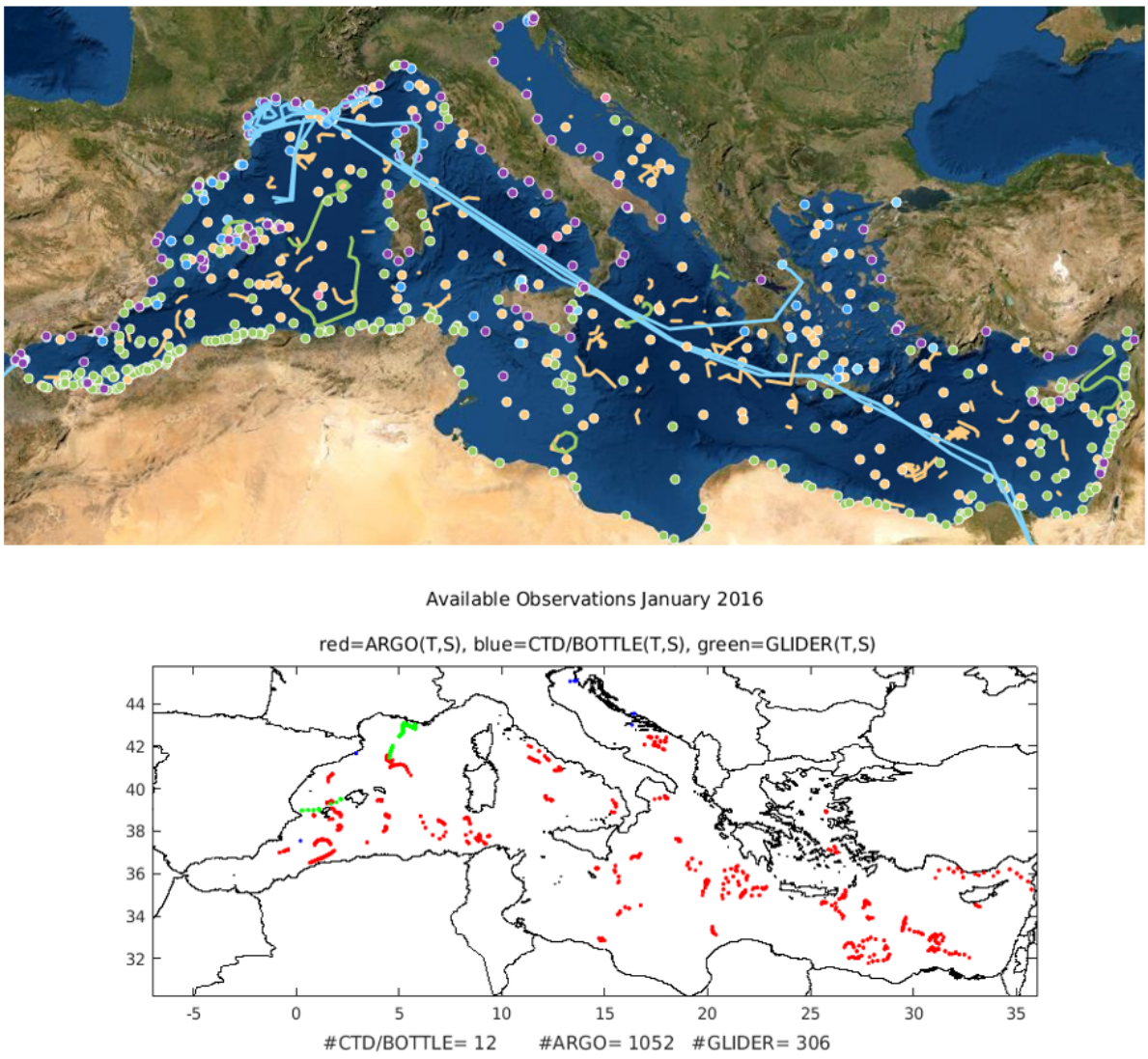

Figure 10 - Location of available (T, S) observations during January 2016 (bottom) and all available in situ oceanographic data for the last 5 years obtained from Copernicus In Situ TAC database (http://www.marineinsitu.eu).

3.3. Application in the SW Iberian Coast

The general methodology described in section 3.1 will be applied in the SW Iberia coast using the SOMA operational modelling system. As explained, SOMA runs in three nested levels, where level zero is a 2D model used to propagate the tide in a larger area and supply it to the boundary of level 1. Levels 1 and 2 are two nested 3D baroclinic models with horizontal resolutions of 2 km and 1 km, respectively. In the implementation of the OSSE, level 2 will be used to provide the NR conditions. While in normal operation the system is forced using forecasted open boundary fields from CMEMS-MERCATOR (GLOBAL_ANALYSIS_FORECAST_PHY_001_024) solution, for these experiments the forcing will be performed using the analysis which replace the forecasts after assimilation in that same prtoduct. The system is restarted every week using also analysed forcing fields. The atmospheric forcing however is kept as a forecast field, as no analysis is available.

The forecast experiments will use level 2. This has the advantage of having a smaller computational cost, while the conclusions are still valid. Since level 2 has a higher resolution, the projection of the

NR space into the forecast space is obtained using a linear operator that computes the average of the NR variables over the level 2 grid. Level 1 and Level 2 grids are co-located, facilitating this procedure.

As referred in section 2.2, the dynamic of SW Iberia coast is characterized by upwelling events, followed by wind relaxation periods. According to Álvarez-Salgado et al (2003), between 66% and 79% of the variability is explained by events with timescales shorter than 30 days. Due to this, a 30-day period will be used for the experiments. Two different experiments will be performed: one during the upwelling season (March to September) in a month with high upwelling activity and another during the non-upwelling season (October to February). During the upwelling season experiment, both upwelling and counter-current events will be analysed for more representativeness.

The EnKF will require the creation of a set of ensemble members. Two different sets will be created, one for each season. The members will be produced perturbing the initial and forcing conditions. A simple random perturbation of those variable generates unbalanced fields not compatible with the system dynamics. This can lead to serious instability problems as pointed out by Quattrocchi et al. (2014). The perturbations will be produced using a method similar to the one described by Turner et al. (2008), where “snapshots” are extracted from the NR and projected into the forecast model space using the projection operator referred above.

The OSSE experiments will use Sea Surface Temperature (SST) and vertical profiles of salinity (S) and temperature (T) as the primary synthetic observations. These values will be extracted from the NR and perturbed using a random function with a normal distribution, and an accuracy equivalent to that existing in the real sensors. Table 2 is used as a guide for sensor accuracies. The synthetic observations used in the Pre-Nautilos and in the Post-Nautilos experiments will be different due to different accuracies and to different datasets. This is particularly relevant in the S and T vertical profiles, where the new technologies developed in Nautilos foresee an increase in the number of buoys and sensors deployed. Figure 19 illustrate the expected scenario of oceanic buoys off the SW Iberian Coast that will be used in the experiments.

Result assessment will be an important component of the experiments. Section 3.1.3 describes different metrics that will be used in the SW Iberian Coast simulations. A detailed evaluation of the relevance and significance of each indicator will be performed.

3.4. APPLICATION IN THE HARDANGER FJORD SYSTEM

The “nature run” for the Hardangerfjord will be simulated using the ROMS 160 m (ROHO160) modelling system described earlier. This new model is an updated version of previous generation 800 m resolution ROMS+ERSEM modelling system for the Hardangerfjord (ROHO800). The coarser resolution grid will be used as the forecasting model for the various OSSE experiments and will assimilate data obtained from the NR as described in 3.1. The older ROHO800 setup will be updated to enable assimilation of data extracted from the “nature run”. Previous simulations, without assimilation, using ROHO800 has been validated and results indicate that the model performed well when we compared physics and biology with observations (Figure 11). Fresh surface waters from river runoff combined with saltwater from the ocean, steep bathymetry, and tides, results in a complex hydrography of the fjord system. Using assimilation of SST, the new improved 160 m model will provide improved physics for the Hardangerfjord. The OSSE approach will identify the best possible locations for where new sensors should be installed in this fjord system so that the next generation biophysical assimilation system will provide enhance modelling results.

The newly developed ROMS+ERSEM modelling system for the Hardangerfjord will be run at a high-resolution of 160 m, which can prove to be challenging in terms of stability of the water masses. We expect to use a very low timestep to be able to correctly spin-up and achieve stable simulations. The steep mountain walls of the fjord require high resolution, non-hydrostatic wind forcing to correctly

represent the atmospheric lateral boundary conditions. The NR will apply the Incremental Strong constraint 4D-Variational (IS4DVAR) system of the Regional Ocean Model System (ROMS) to assimilate SST (Moore et al. 2009). The period of simulation will be 2017-2019 (2.5-3 years) and include SST from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system that generates global, daily, gap-filled foundation sea surface temperature (SST) fields from satellite data and in situ observations (Good et al. 2020) (See Figure 23).

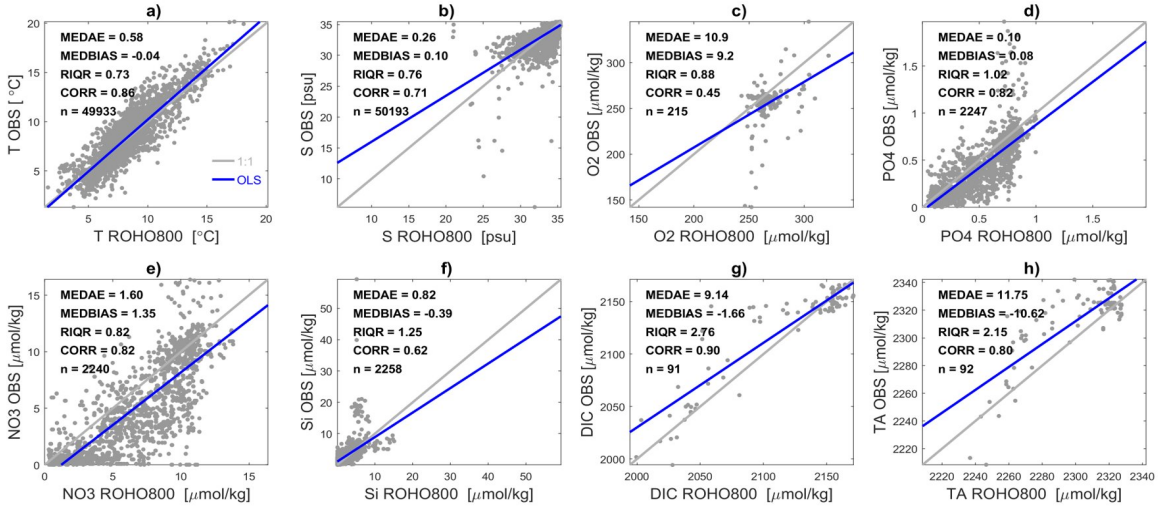

Figure 11 - ROHO800 model results of physics and biology compared with observations; a) temperature, b) salinity, c) oxygen d) phosphate, e) nitrate, f) silicate, g) dissolved inorganic carbon, h) total alkalinity.

The OSSE experiments will use the vertical profiles of salinity and temperature (including sea surface temperature) as observations, which will be extracted from the NR and used to perturb the initial conditions.

4. PRODUCED DATASETS

4.1. DATASETS PRODUCED IN THE MEDITERRANEAN

In Figure 12, the model skill of the nature run with assimilation (NR) and without assimilation (FR) is evaluated against altimetry data. The example illustrates the impact of data assimilation on the simulated sea surface height (SSH), which is representative of near surface circulation patterns (i.e. low SSH=cyclone, high SSH=anti-cyclone). The simulated circulation without assimilation, as depicted by the SSH seasonal patterns, reproduces reasonably well most of the major features (see Figure 2), characterizing the Mediterranean near surface circulation, such as the semi-permanent anti-cyclonic (Gulf of Syrte, Alboran Sea) and cyclonic (Gulf of Lions, Southern Adriatic) features, major currents (Algerian, Liguro-Provencal, Asia Minor, Atlantic-Ionian) and less permanent features (Ierapetra anti-cyclone). Some features, such as Rhodes and Mersa-Matruh gyres are partly reproduced. The simulated circulation is significantly improved when data assimilation is employed, as indicated in the Taylor diagram (Figure 13), showing an increase of correlation (from ~0.6 to ~0.9) and decrease of RMSD (from ~0.8 to ~0.4) of the model SSH against satellite altimetry data. The simulated SST fields (not shown) also present a good agreement with satellite data, showing a correlation above 0.9. The model score of FR and NR with regard to different metrics (see section 3.1.3) is shown in Table 3.

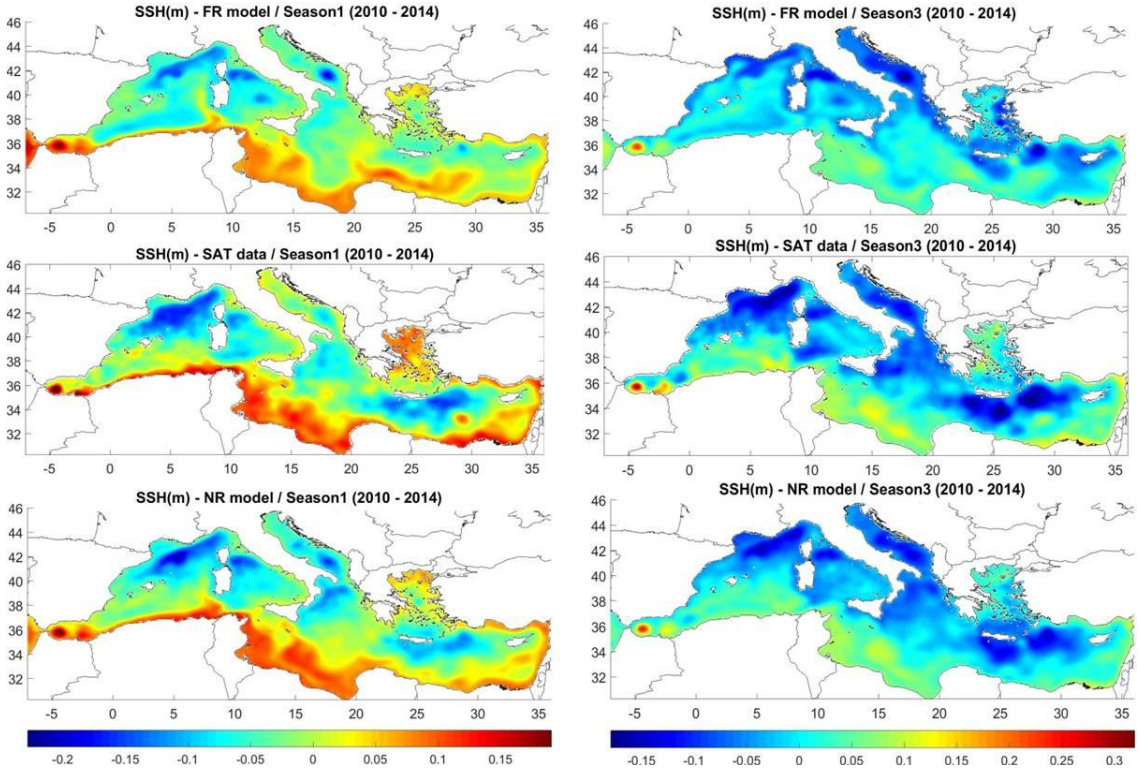

Figure 12 - Comparison of the seasonal (1=winter, 2=spring, 3=summer, 4=autumn) climatology of sea surface height (SSH) as derived with (left, Free Run) and without (right, Natural Run) data assimilation run of the model for the years 2010-2014 against AVISO+ altimetry data (middle).

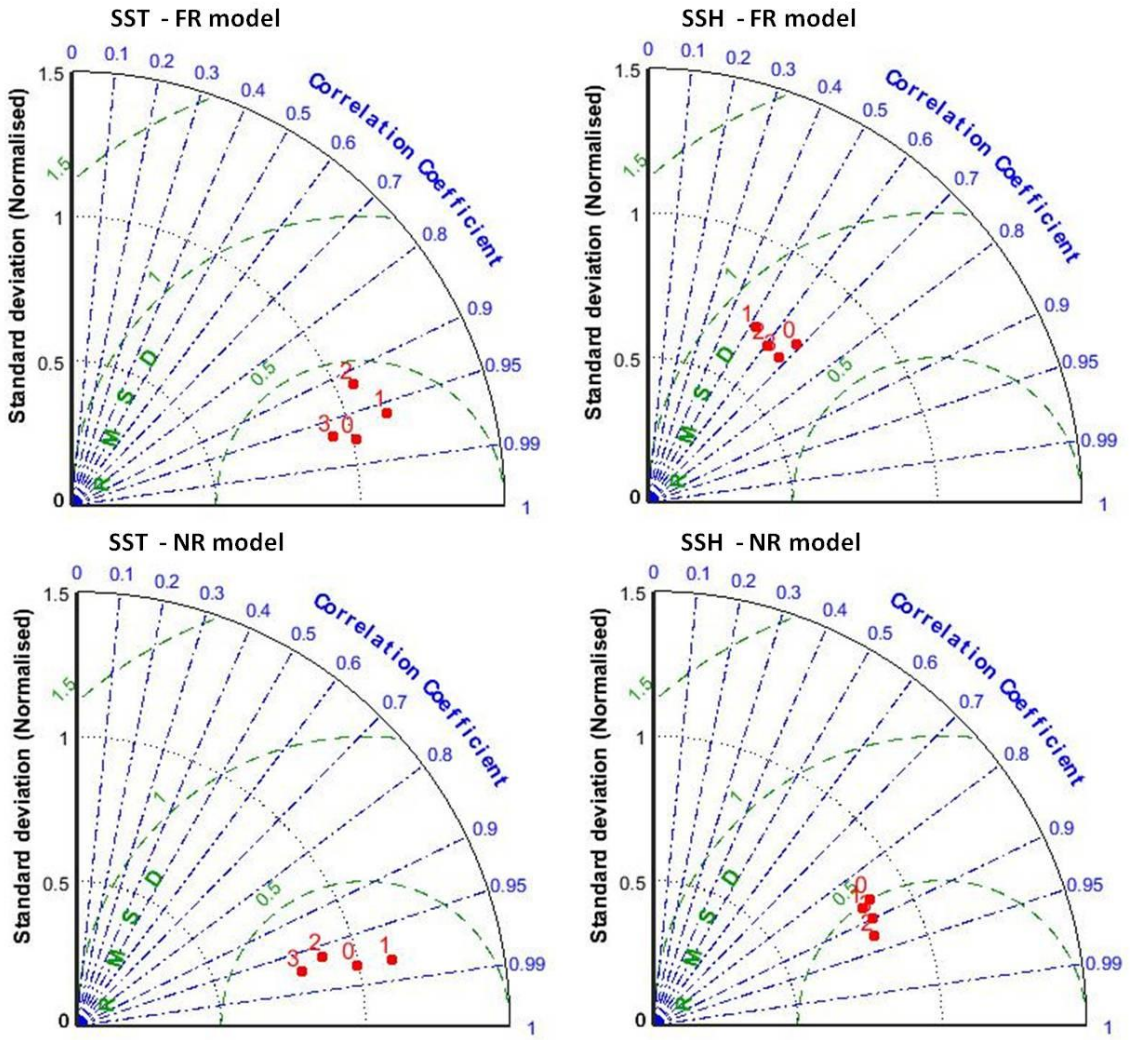

Figure 13 - Taylor diagram of mean (2010-2014) seasonal (0=winter, 1=spring, 2=summer, 3=autumn) simulated sea surface height (SSH) and temperature (SST), with (up, FR) and without data assimilation (right, HR), against satellite data over the same period from AVISO+ and Copernicus data base, respectively.

Table 3 - Percentage bias (PBIAS=(AVG_{model}–AVG_{data})/AVG_{data}), normalized standard deviation (STDN=STD_{model}/STD_{data}), root mean square deviation (RMSD), Pearson correlation coefficient (ρ) and the Murphy skill score (Σ) of free (FR) and natural (NR) runs of the model simulated against available satellite data over 2010–2014 period.

VARIABLE	SEASON	MODEL	PBIAS	STDN	RMSD	ρ	Σ
SSH (m)	1	FR	118.0	0.75	0.05	0.68	0.32
		NR	42.9	0.93	0.03	0.90	1.65
	2	FR	84.0	0.71	0.05	0.52	-0.02
		NR	29.3	0.84	0.03	0.89	1.66
	3	FR	-158.7	0.68	0.05	0.61	-0.61
		NR	-58.9	0.83	0.03	0.92	1.59
	4	FR	-38.9	0.68	0.05	0.67	-2.08
		NR	-15.3	0.91	0.03	0.89	1.22
SST (°C)	1	FR	5.3	0.96	0.35	0.97	1.66
		NR	5.3	0.96	0.36	0.97	1.65
	2	FR	2.1	1.14	0.36	0.96	1.75
		NR	1.2	1.04	0.24	0.98	1.91
	3	FR	1.2	1.06	0.55	0.92	1.76
		NR	0.9	0.98	0.27	0.98	1.92
	4	FR	2.3	0.93	0.46	0.97	1.86
		NR	1.2	0.96	0.39	0.98	1.94

The improvement of the model simulated SSH with data assimilation may be also seen with the SSH RMSD evolution with time (Figure 14), showing a constantly lower error (x 1/3), as compared to the free run without assimilation.

Finally, the horizontal variability of the SSH (relative difference) and SST (mean difference) model deviation from satellite data, with and without assimilation is depicted in Figure 15, The NR SST shows a significantly reduced bias, especially in coastal areas. An even stronger decrease is also depicted in the decrease of the SSH relative error, being slightly lower in the Aegean Sea.

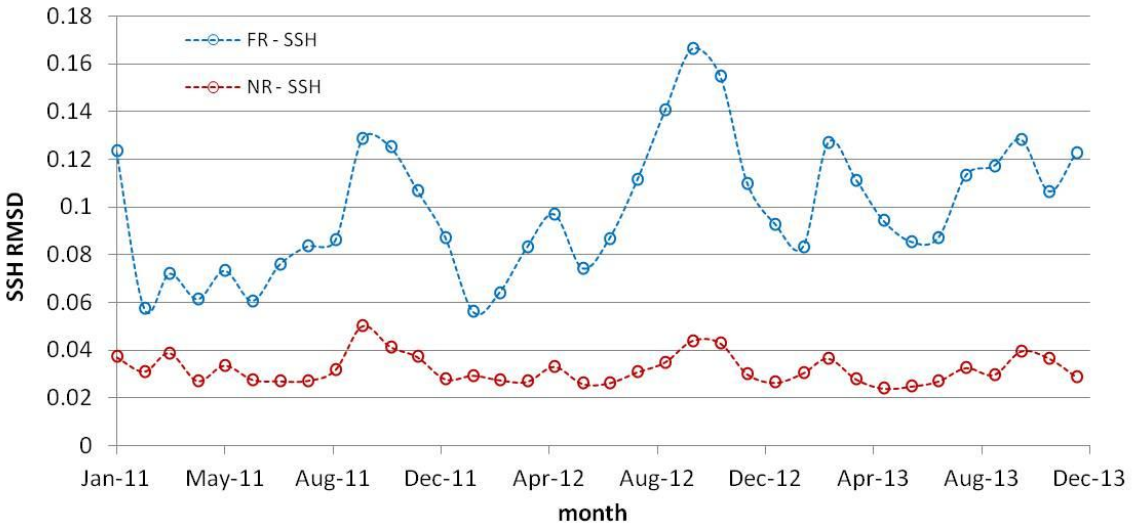

Figure 14 - Spatially integrated root mean square deviation (RMSD, m) between the forecasting (free and natural) model results and satellite observations over 2011–2013.

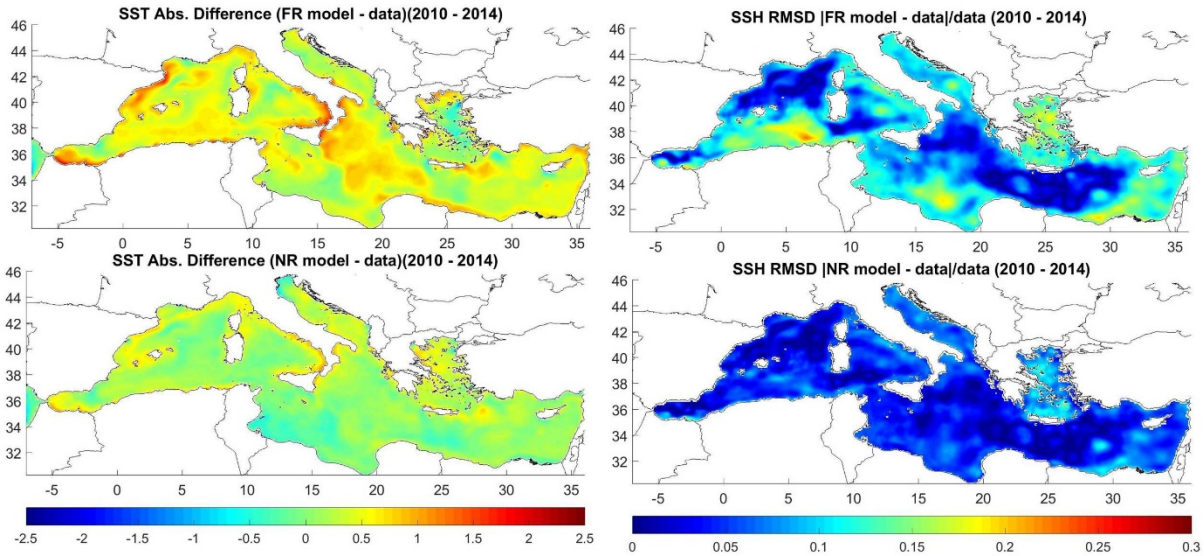

Figure 15 - Model simulation of average sea surface temperature (SST) difference (left) and sea surface height (SSH) absolute relative error (right), without (FR) and with (NR) assimilation scheme during 2010–2014.

4.2. DATASETS PRODUCED IN THE SW IBERIAN COAST

4.2.1. Nature Run and Free Run

The first experiment for the SW Iberian coast was conducted during the upwelling season. Consequently, the simulations, carried out for one month, spanned from April 2nd to the same day in May 2021. As indicated in section 2.2, SOMA forecast solution is restarted from the conditions of CMEMS global model once a week. Therefore, it was determined that the Nature Run (NR), as the "hypothetical truth", would be the results of the forecasts, within the indicated period, with weekly restarts and the Free Run (FR), without restart. Other variability imposed between the runs was that, for FR, only level 1 of SOMA was used (2 km grid), so that its results would have a lower resolution in relation to the NR.

The experiment period was chosen so that the two most important phenomena of the Algarve coast are present. The upwelling was present in the beginning of the month, losing strength along time and the countercurrent settles down close to the end of the month. Figure 16 presents snapshots of these moments for NR and Figure 17, for FR. All analyzes and comparisons between NR and FR were performed relative to the geographic limits of level 2. The similarity in A1 of Figure 16 and Figure 17 is apparent, but as NR goes through weekly restarts to adjust ocean variables with CMEMS conditions, FR solutions diverges from NR's, as indicated by B1 in the figures. The evolution of this divergence is presented in Figure 18.

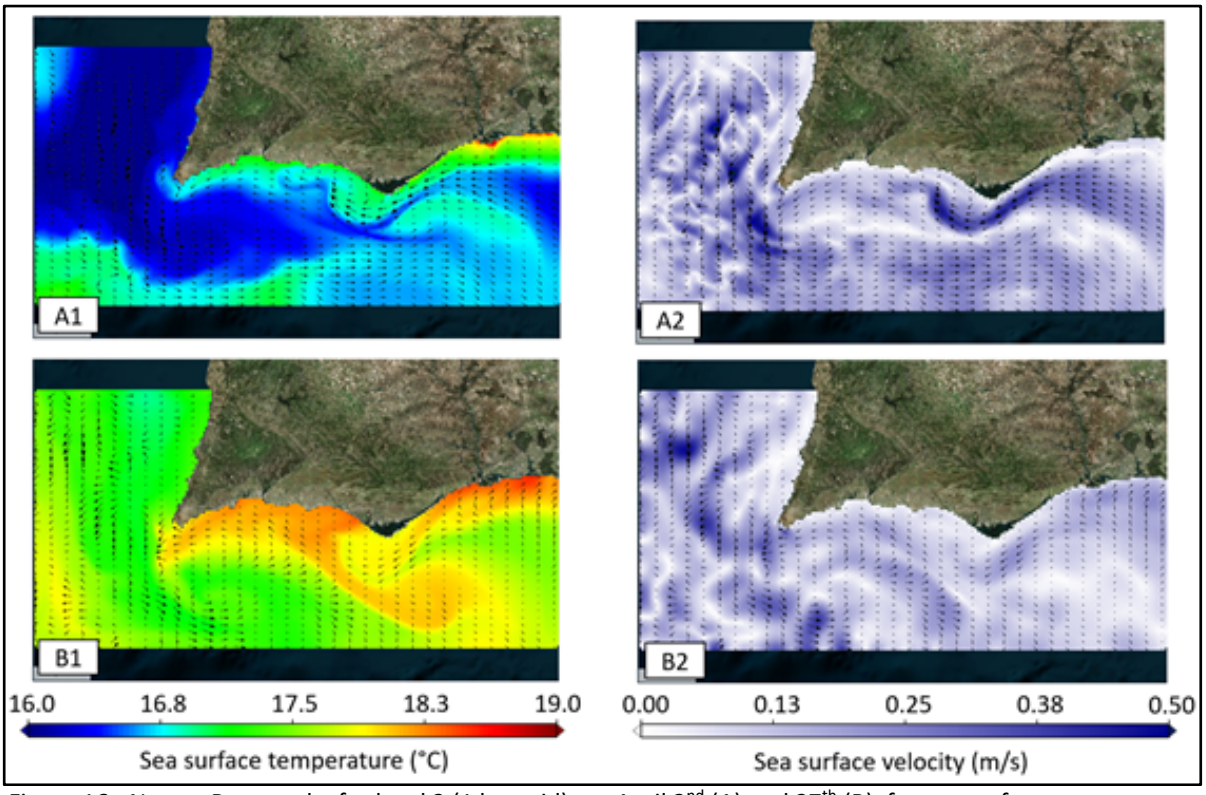

Figure 16 - Nature Run results for level 2 (1 km grid), on April 2nd (A) and 27th (B), for sea surface temperature (1) and velocity (2).

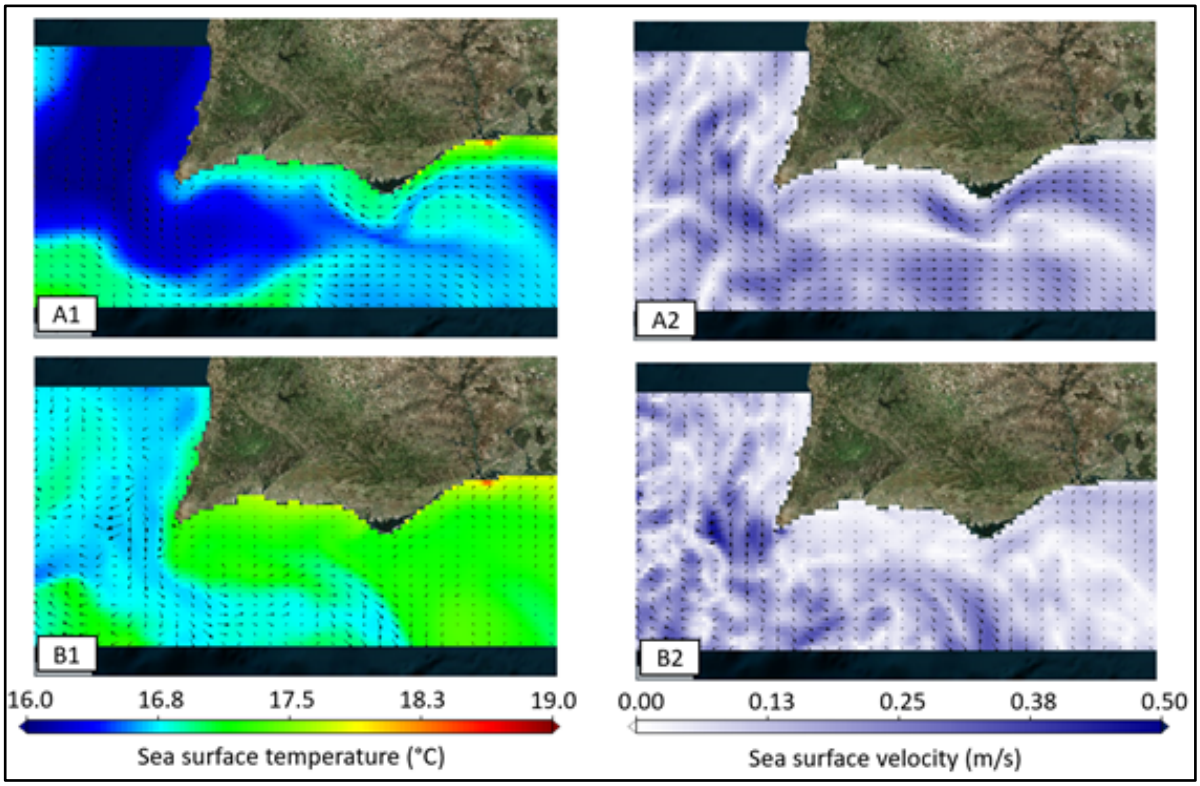

Figure 17 - Free Run results for level 1 (2 km grid), cut in level 2 domain, on April 2nd (A) and 27th (B), for sea surface temperature (1) and velocity (2).

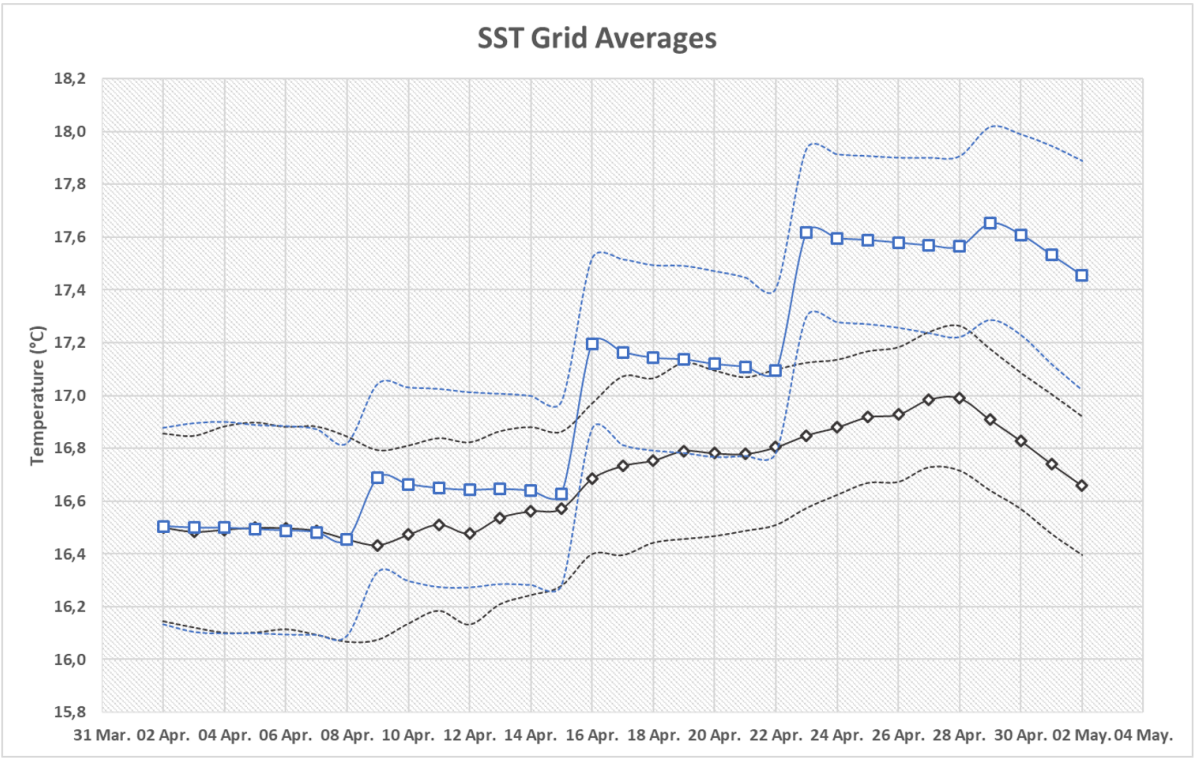

Figure 18 - Daily averages of SST for the NR level 2 (blue line), and of SST for the FR level 1 (black line). Dashed lines are the standard deviation for each curve. The blue line jumps correspond to NR restarts. At each jump the error between NR and FR increases.

4.2.2. *Pre-NAUTILOS and Post-NAUTILOS synthetic observation datasets*

The data assimilation system for SOMA is still under development. Nevertheless, synthetic observations were extracted from the Nature Run simulations, following the procedure to assimilate Free Run results. Currently, there are only four buoys, belonging to the Instituto Hidrográfico (IH) of Portugal Navy, within the limits of SOMA domain, namely: Coastal Faro, Oceanic Faro, Coastal Sines and Oceanic Sines. However, only Coastal Faro is located inside of the Algarve Coastal model highest resolution level and, yet, it is not operational at the moment. In the context of NAUTILOS we assume Coastal Faro buoy as the single buoy available for the Pre-NAUTILOS OSSE. For the Post-NAUTILOS scenario five other buoys are assumed inside SOMA level 2 domain. Table 4 provide the locations of each synthetic observation spot, from where the synthetic observations will be extracted. Figure 19 show the location of these new "virtual" buoys and Figures Figure 20 to Figure 21, the temperature and salinity profiles of the NR extracted from them.

Table 4 - Synthetic observations locations.

BUOY	LATITUDE	LONGITUDE
IH Coastal Faro	36.905	-7.898
Pre-Nautilos buoy #1 (PN_1)	37.195	-9.200
Pre-Nautilos buoy #2 (PN_2)	36.805	-9.405
Pre-Nautilos buoy #3 (PN_3)	36.795	-8.550
Pre-Nautilos buoy #4 (PN_4)	36.705	-7.895
Pre-Nautilos buoy #5 (PN_5)	36.745	-7.305

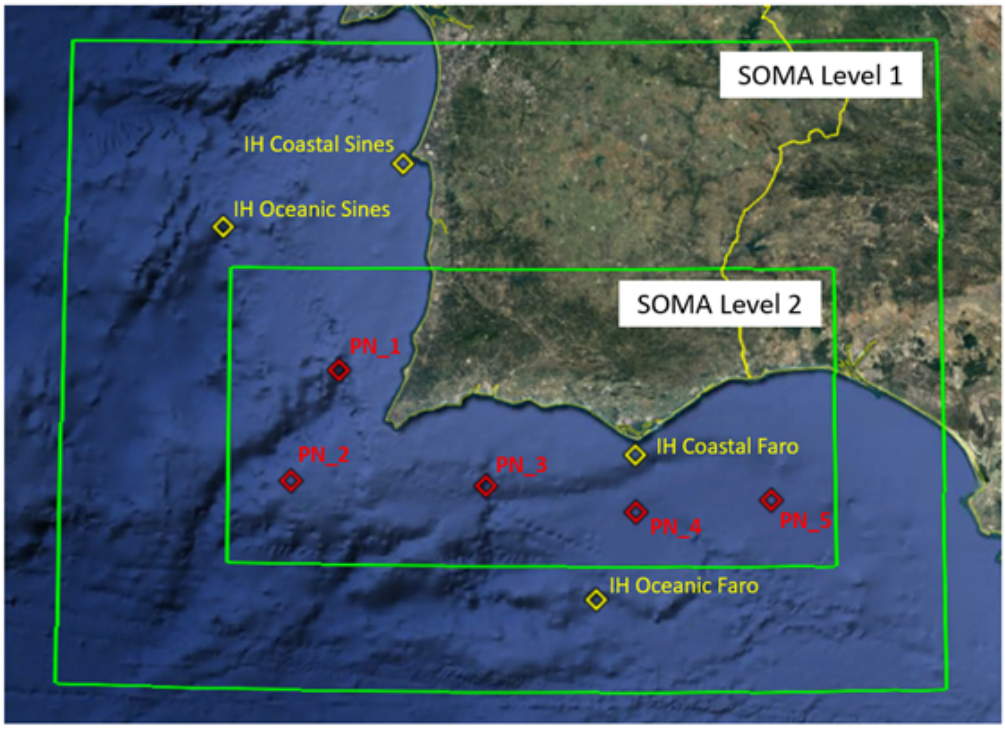

Figure 19 - Map of south Portugal with the indications of SOMA domains limits, the location of the buoys from Instituto Hidrográfico and the five new spots for synthetic observations.

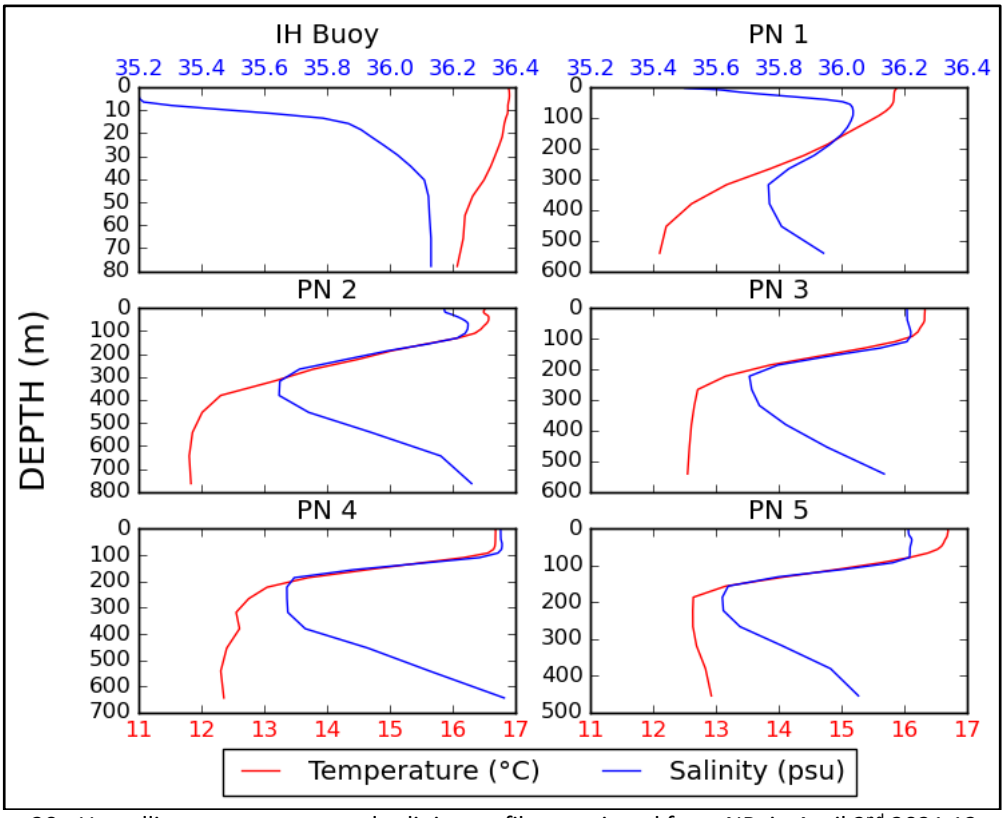

Figure 20 - Upwelling temperature and salinity profiles, retrieved from NR, in April 2nd 2021 12pm of simulation time, at the locations of the buoys defined in Table 4.

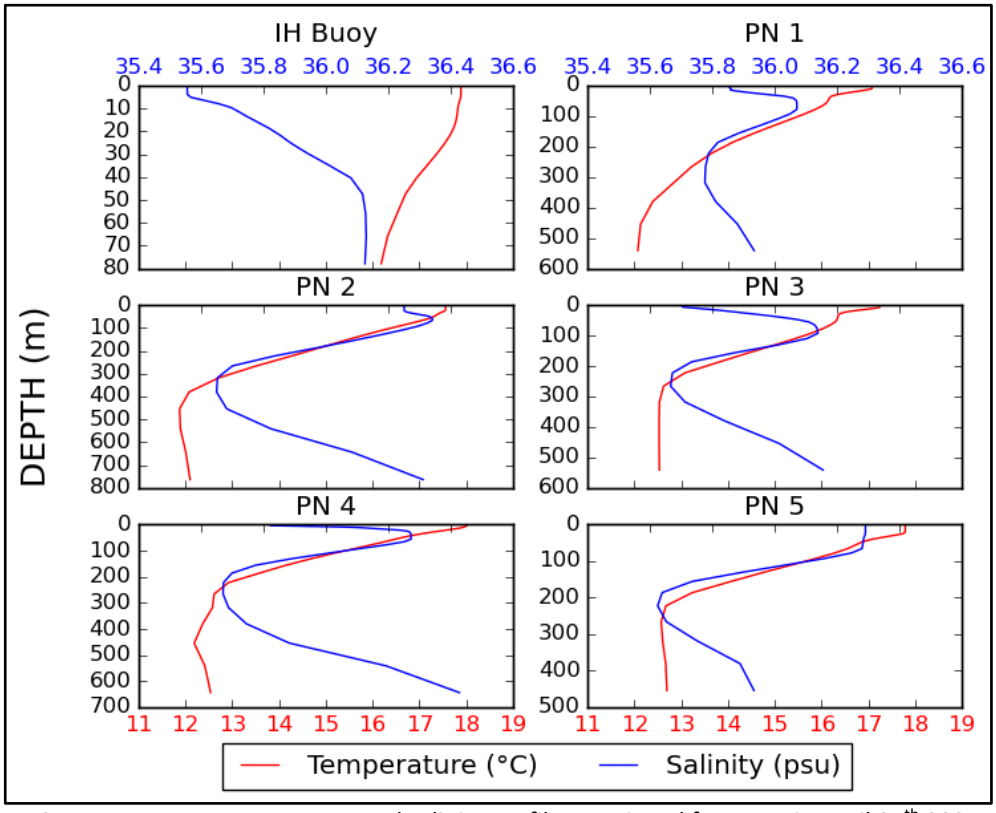

Figure 21 - Counter current temperature and salinity profiles, retrieved from NR, in April 27th 2021 12pm of simulation time, at the locations of the buoys defined in Table 4.

4.3. DATASETS PRODUCED IN THE HARDANGER FJORD SYSTEM

The first initial testing of the model has been conducted although for a very short time-period (Figure 22). The first simulations were performed using assimilation of OSTIA SST and GLORYS12V1 as ocean boundary forcing. The NR is run with restart of initial conditions based on assimilating OSTIA SST (Figure 23) as input every 3 days and then restarting the simulation, adjusting the solution by iterating back and forth until simulation is close to the observations, then repeat the procedure for the next 3 days (IS4DVAR assimilation). The current dataset produced only covers one week NR.

The experiment for Hardangerfjord will focus on the spring period of 3 years 2017-2019 where we will simulate the release of freshwater from snow melting in the mountains entering the fjord system as freshwater, and mixing with the water masses of the fjord to produce a gradually stratifying environment triggering spring bloom conditions.

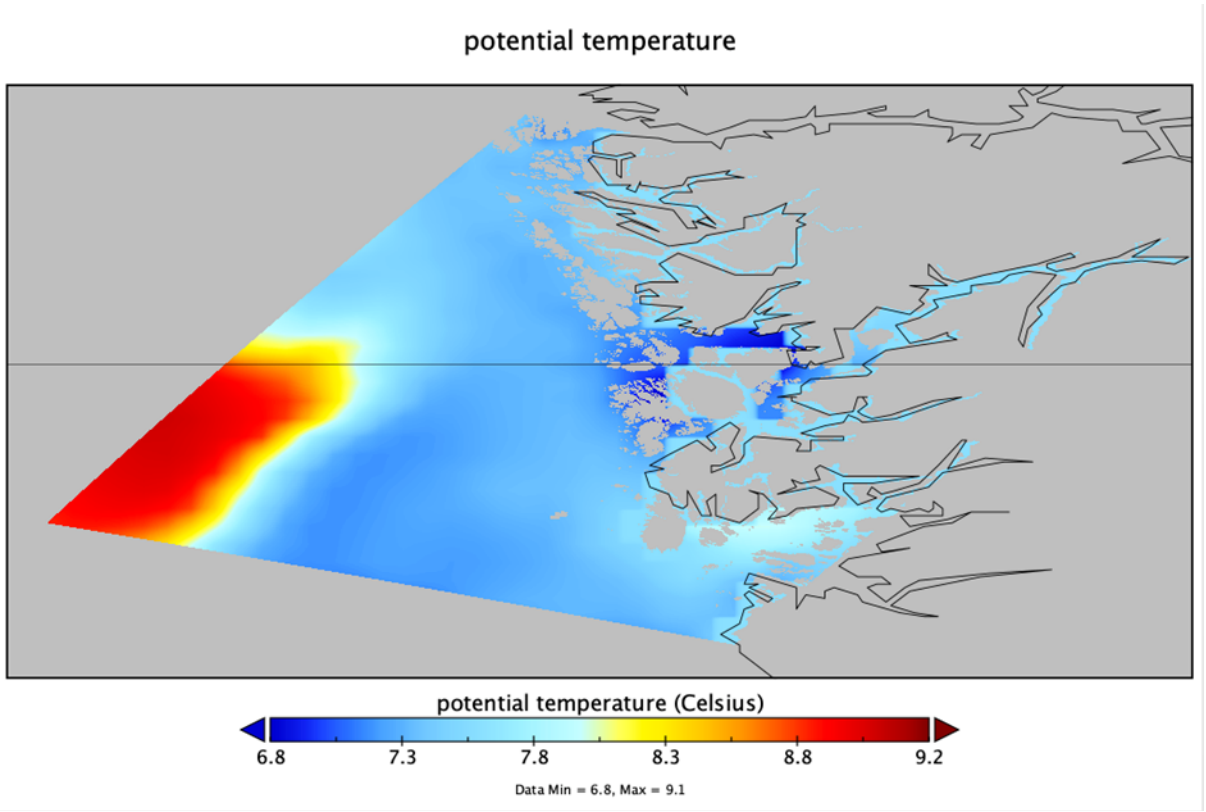

Figure 22 - Showing the initial conditions for the ROMS simulation starting on January 1st 2017.

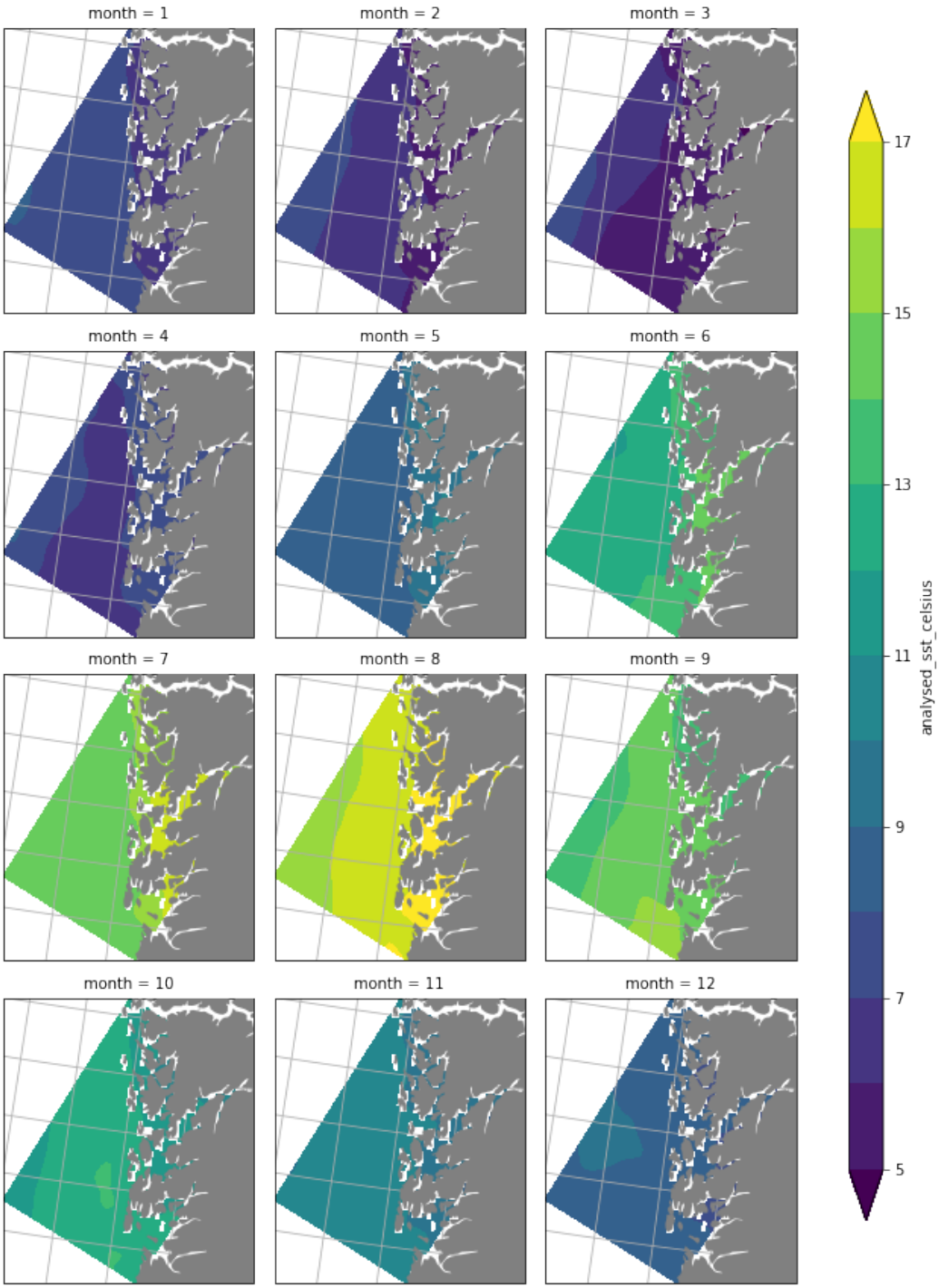

Figure 23 - Monthly climatology of OSTIA SST for the period 2017-2019 to be assimilated into the ROHO160 model as part of NAUTILOS.

Several monitoring stations are located within the Hardangerfjord as part of a national Norwegian observing system to monitor the health of the coastal oceans. Our locations for pre-NAUTILOS subsampling will overlap with the locations of stations VT69, VT70, VT74, and VT53 to be able to compare directly with observations (Figure 24).

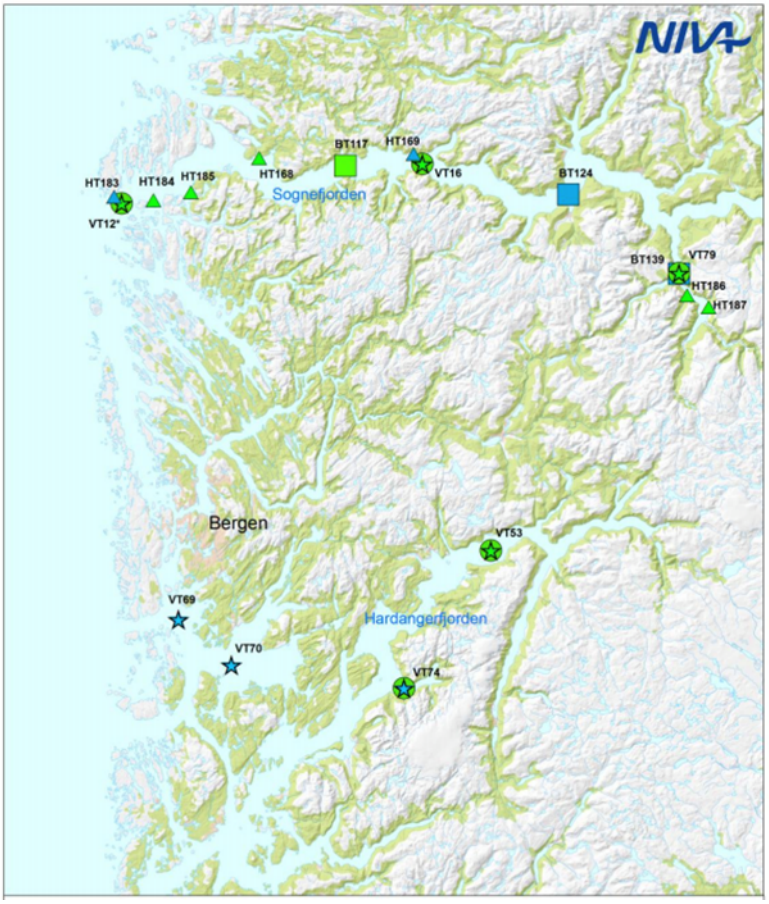

Figure 24 - Monitoring stations within the Hardangerfjord as part of a national Norwegian observing system. Pre-NAUTILOS subsampling overlap stations VT69, VT70, VT74, and VT53 to allow direct comparison with observations.

The locations of the post-NAUTILOS observation datasets have not been decided yet but will be at least twice the number of current observations (4) and will be positioned covering the entire length of the Hardangerfjord. The post-NAUTILOS observing data will have uncertainty identical to the sensors they are representing, and the synthetic observations will be extracted from the "nature run" simulations, following the procedure to assimilate Free Run results.

5. Discussion and Conclusions

In this deliverable the conceptual basis for the execution of the simulation experiments in NAUTILOS are defined. Their objective is to evaluate the impact of the new observation technologies on model accuracy and forecasting capabilities. The experiments itself, will be conducted in WP9, following this deliverable. The experiments will take into account the characteristics of the sensors and platforms being defined in NAUTILOS as well as the specificalities of each study site. Since three study sites are addressed, with different dynamics, using different modelling suits, and addressing different processes, the conclusions provided by the simulation experiments will be sufficiently generic to consider them representative of other regions.

Despite their differences, the simulation experiments will share a common basis, using the concept of Observing System Simulation Experiments (OSSE) where a high-resolution simulation of the system, named Nature Run (NR) is used to emulate the true state. Synthetic observations will then be extracted from this NR to emulate the real observations. Two sets of synthetic observations will be created, considering the Pre-Nautilos and Post-Nautilos situations. These take into consideration the number and location of the observations as well as their accuracy in both situations.

The methods to compare the experiment results with the NR are essential to this framework. Several standard methods common in the literature are proposed, enabling a comprehensive evaluation of the model behaviour.

This deliverable also reports the preliminary model executions in preparation for the simulations in WP9. The scenarios dates and lengths for the experiments were defined for each site, and the NR where executed. Preliminary evaluation of NR adequacy was performed to assure the NR characteristics are representative of the system dynamics.

As referred, the work reported in this deliverable is the preparatory phase of the simulation experiments that will be developed in WP9. In WP9 the synthetic observations will be extracted from the NR, the simulation scenarios will be executed, and the behaviour of the models will be analysed, to assess the improvement in model accuracy and forecasting skill produced by the NAUTILOS new technologies.

6. REFERENCES

Álvarez, I., Castro, M., Prego, R., & Gómez-Gesteira, M. (2003). Hydrographic characterization of a winter-upwelling event in the Ria of Pontevedra (NW Spain). Estuarine, Coastal and Shelf Science, 56(3–4), 869–876. https://doi.org/10.1016/S0272-7714(02)00309-8

Álvarez-Salgado, X.A., Figueiras, F.G., Pérez, F.F., Groom, S., Nogueira, E., Borges, A.V., Chou, V., Castro, C.G., Moncoiffé, G., Ríos, A.F., Miller, A.E.J., Frankignoulle, M., Savidge, G. and Wollast, R., 2003. The Portugal coastal counter current off NW Spain: new insights on its biogeochemical variability. Progress in Oceanography, 56, 2, 281-321.

Aníbal, J., Gomes, A., Mendes, I., & Moura, D. (2019). Ria Formosa: Challenges of a coastal lagoon in a changing environment. Universidade do Algarve Editora.

Atlas, R., Kalnay, E. and Halem, M., 1985. Impact of satellite temperature soundings and wind data on numerical weather prediction. Optical Engineering, 24, 2, 341–346.

Atlas, R., 1997. Atmospheric observations and experiments to assess their usefulness in data assimilation. J. Meteor. Soc. Japan, 75, 111–130.

Baretta, J.W., Ebenhoh, W. and Ruardij, P., 1995. The European regional seas ecosystem model, a complex marine ecosystem model. Netherlands Journal of Sea Research, 33, 233-246.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific Data, 5(1), 180214. https://doi.org/10.1038/sdata.2018.214

Blumberg, A. and Mellor, G., 1987. A description of a three-dimensional coastal ocean circulation model. In Three-Dimensional Coastal Ocean Models, N.S. Heaps (Ed.). AGU publications, Wiley.

Blumberg, A. F., & Kantha, L. H. (1985). Open Boundary Condition for Circulation Models. Journal of Hydraulic Engineering, 111(2), 237–255. https://doi.org/10.1061/(ASCE)0733-9429(1985)111:2(237)

Bosc, E., Bricaud, A. et Antoine, D., 2004. Seasonal and interannual variability in algal biomass and primary production in the Mediterranean Sea, as derived from 4 years of SeaWiFS observations, Global Biogeochem Cycles, 18 (GB1005).

Brankart, J.-M. and Brasseur, P., 1998. The general circulation in the Mediterranean Sea: A climatological approach. Journal of Marine Systems, 18, 41-70.

Brasseur, P., Beckers, J. M., Brankart, J. M. and Schoenauen, R.,1996. Seasonal temperature and salinity fields in the Mediterranean Sea: climatological analyses of an historical data set. Deep Sea Research., 43, 159–192.

Braunschweig, F., Leitao, P. C., Fernandes, L., Pina, P., & Neves, R. J. J. (2004). The object-oriented design of the integrated water modelling system MOHID. In Developments in Water Science (Vol. 55, pp. 1079–1090). Elsevier. https://doi.org/10.1016/S0167-5648(04)80126-6

Broquet, G., Moore, A. M., Arango, H. G., Edwards, C. A., & Powell, B. S. (2009). Ocean state and surface forcing correction using the ROMS-IS4DVAR data assimilation system. Mercator Ocean Quarterly Newsletter, 34, 5-13.

Bruggeman, Jorn, and Karsten Bolding. 2014. “A General Framework for Aquatic Biogeochemical Models.” Environmental Modelling & Software 61 (November): 249–65.

Bryden, H.L. and Kinder, T.H., 1991. Steady two-layer exchange through the Strait of Gibraltar. Deep-Sea Research, 38, S445-S463.

Buongiorno Nardelli, B., Tronconi, C., Pisano, A., and Santoleri, R., 2013. High and Ultra-High resolution processing of satellite Sea Surface Temperature data over Southern European Seas in the framework of MyOcean project. Remote Sensensing and Environment, 129, 1– 16.

Butenschön, Momme, James Clark, John N. Aldridge, Julian Icarus Allen, Yuri Artioli, Jeremy Blackford, Jorn Bruggeman, et al. 2016. "ERSEM 15.06: A Generic Model for Marine Biogeochemistry and the Ecosystem Dynamics of the Lower Trophic Levels." Geoscientific Model Development 9 (4): 1293–1339.

Costa, M., Silva, R., & Vitorino, J. (2001). Contribuição para o estudo do clima de agitação marítima na costa portuguesa. 2as Jornadas Portuguesas de Engenharia Costeira e Portuária, 20.

Crise, A., Allen, J.I., Baretta, J., Crispi, G., Mosetti, R. and Solidoro, C., 1999. The Mediterranean pelagic ecosystem response to physical forcing. Progress in Oceanography, 44, 219–243.

Dias, L. F., Aparício, B. A., Nunes, J. P., Morais, I., Fonseca, A. L., Pastor, A. V., & Santos, F. D. (2020). Integrating a hydrological model into regional water policies: Co-creation of climate change dynamic adaptive policy pathways for water resources in southern Portugal. Environmental Science & Policy, 114, 519–532. https://doi.org/10.1016/j.envsci.2020.09.020

Fiúza, A. F. G. (1983). Upwelling Patterns off Portugal. In E. Suess & J. Thiede (Eds.), Coastal Upwelling Its Sediment Record (pp. 85–98). Springer US. https://doi.org/10.1007/978-1-4615-6651-9_5

Fiúza, de Macedo, & Guerreiro. (1982). Climatological space and time-variation of the Portuguese coastal upwelling. Oceanologica Acta, 5(1): 31-40.

Flather, R.A., 1976. A tidal model of the northwest European continental shelf. Memoires Societe Royale des Sciences de Liege, 6, 141–164

Frouin, R., Fiúza, A. F. G., Ambar, I., & Boyd, T. J. (1990). Observations of a poleward surface current off the coasts of Portugal and Spain during winter. Journal of Geophysical Research, 95(C1), 679. https://doi.org/10.1029/JC095iC01p00679

Garel, E., Laiz, I., Drago, T., & Relvas, P. (2016). Characterisation of coastal counter-currents on the inner shelf of the Gulf of Cadiz. Journal of Marine Systems, 155, 19–34. https://doi.org/10.1016/j.jmarsys.2015.11.001

Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J., & Worsfold, M. (2020). The Current Configuration of the OSTIA System for Operational Production of Foundation Sea Surface Temperature and Ice Concentration Analyses. Remote Sensing, 12(4), 720. https://doi.org/10.3390/rs12040720

Gunther, H. and Behrens, A. 2012. The WAM model. Validation document Version 4.5.4, Intitute of Coastal Research Helmholtz-Zentrum Geesthach.

Halliwell Jr., G. R., Kourafalou, V., Le Hénaff, M., Shay, L. K. and Atlas, R., 2015. OSSE impact analysis of airborne ocean surveys for improving upper-ocean dynamical and thermodynamical forecasts in the Gulf of Mexico. Progress in Oceanography, 130, 32–46.

Halliwell Jr., G. 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, 105–130.

Haynes, R., & Barton, E. D. (1990). A poleward flow along the Atlantic coast of the Iberian peninsula. Journal of Geophysical Research, 95(C7), 11425. https://doi.org/10.1029/JC095iC07p11425

Hoteit, I., Pham, D.T., Blum J., 2001. A semi-evolutive partially local filter for data assimilation. Marine Pollution Bulletin 43, 64–174.

Hoteit, I., X. Luo, X. and Pham, D.-T., 2012. Particle kalman filtering: A nonlinear bayesian framework for ensemble kalman filters. Monthly Weather Review, 140/2, 528-542.

Husa, Vivian, Tina Kutti, Arne Ervik, Kjersti Sjøtun, Pia Kupka Hansen, and Jan Aure. 2014. "Regional Impact from Fin-Fish Farming in an Intensive Production Area (Hardangerfjord, Norway)." Marine Biology Research 10 (3): 241–52.

IH. (2009). Instituto Hidrográfico—Tabela de Marés 2010 (Vol. 1). Ministério da Defesa Nacional, Marinha. https://www.hidrografico.pt/tabelamares

IP, M. (2006). List of internal metrics for the MERSEA-GODAE Global Ocean: Specification for implementation.

Janeiro, J., Martins, F., & Relvas, P. (2012). Towards the development of an operational tool for oil spills management in the Algarve coast. Journal of Coastal Conservation, 16(4), 449–460. https://doi.org/10.1007/s11852-012-0201-8

Janeiro, J., Neves, A., Martins, F., & Relvas, P. (2017). Integrating technologies for oil spill response in the SW Iberian coast. Journal of Marine Systems, 173, 31–42. https://doi.org/10.1016/j.jmarsys.2017.04.005

Janeiro, J., Zacharioudaki, A., Sarhadi, E., Neves, A., & Martins, F. (2014). Enhancing the management response to oil spills in the Tuscany Archipelago through operational modelling. Marine Pollution Bulletin, 85(2), 574–589. https://doi.org/10.1016/j.marpolbul.2014.03.021

Kalaroni S., Tsiaras K., Petihakis G., Economou-Amilli A., Triantafyllou G., 2020. Modelling the Mediterranean Pelagic Ecosystem using the POSEIDON ecological model. Part II: Biological dynamics, Deep Sea Research Part II: Topical Studies in Oceanography, Volume 171, 2020, 104711, ISSN 0967-0645, https://doi.org/10.1016/j.dsr2.2019.104711.

Kallos, G., Nickovic, S., Papadopoulos, A., Jovic, D., Kakaliagou, O., Misirlis, N., & Manousakis, M. (1997). The regional weather forecasting system SKIRON: An overview. Proceedings of the Symposium on Regional Weather Prediction on Parallel Computer Environments, 15, 17.

Korres, G. and Lascaratos, A., 2003. A one-way nested eddy resolving model of the Aegean and Levantine basins: implementation and climatological runs. Annales Geophysicae, 21, 205–220.

Korres, G., Hoteit, I., Triantafyllou, G., 2007. Data assimilation into a Princeton Ocean Model of the Mediterranean Sea using advanced Kalman filters, Journal of Marine systems, 65, 84-104, doi.org/10.1016/j.jmarsys.2006.09.005.

Kourafalou, V. and Tsiaras, K., 2007. A nested circulation model for the North Aegean Sea. Ocean Science, (April), 343–372.

Lascaratos, A., Roether, W., Nittis, K. and Klein, B., 1999. Recent changes in deep water formation and spreading in the eastern Mediterranean Sea: a review. Progress in Oceanography 44, 5-36.

Leitão, P., Coelho, H., Santos, A., & Neves, R. (2005). Modelling the main features of the Algarve coastal circulation during July 2004: A downscaling approach. Journal of Atmospheric & Ocean Science, 10(4), 421–462. https://doi.org/10.1080/17417530601127704

Lellouche, Jean-Michel, Olivier Le Galloudec, Eric Greiner, Gilles Garric, Charly Regnier, Marie Drevillon, Romain Bourdallé-Badie, Clément Bricaud, Yann Drillet, and Pierre-Yves Le Traon. 2018. "The Copernicus Marine Environment Monitoring Service Global Ocean 1/12° Physical Reanalysis GLORYS12V1: Description and Quality Assessment." In , 19806.

Liu, Y., and Weisberg, R. H., 2011. Evaluation of trajectory modeling in different dynamic regions using normalized cumulative Lagrangian separation. J. Geophys. Res., 116, C09013, doi:10.1029/2010JC006837.

Lorente, P., Aznar, R., Alvarez Fanjul, E., Pascual, Á., Toledano Lozano, C., Amo, A., Dabrowski, T., Levier, B., Reffray, G., Dalphinet, A., & Aouf, L. (2019). The NARVAL software toolbox in support of ocean model skill assessment at regional and coastal scales.

Lorenz, E. N., 1956. Empirical orthogonal functions and statistical weather prediction. Statistical Forecasting Project Rep. 1, MIT Department of Meteorology, 49 pp.

Martins, F. (1999). Modelação matemática tridimensional de escoamentos costeiros e estuarinos usando uma abordagem de coordenada vertical genérica [PhD Dissertation]. Universidade Técnica de Lisboa.

Martins, F., Leitão, P., Silva, A., & Neves, R. (2001). 3D modelling in the Sado estuary using a new generic vertical discretization approach. Oceanologica Acta, 24, 51–62. https://doi.org/10.1016/S0399-1784(01)00092-5

Martins, F., Neves, R. J., & Leitão, P. C. (1998). A three-dimensional hydrodynamic model with generic vertical coordinate.

Martinsen, E. A., & Engedahl, H. (1987). Implementation and testing of a lateral boundary scheme as an open boundary condition in a barotropic ocean model. Coastal Engineering, 11(5–6), 603–627. https://doi.org/10.1016/0378-3839(87)90028-7

Mellor, G. L. and Yamada, T., 1982. Development of a turbulence closure model for geophysical fluid problems. Reviews of Geophysics, 20(4), 851.

Mellor, G. L., 2004. Users guide for ocean model. Ocean Modelling, 8544, 0710.

Mendonça, F. (2020). A Python Based System to Manage Simulations of Coastal Operational Models [Msc Dissertation]. Universidade do Algarve.

Menna, M., Poulain, P.-M., Zodiatis, G., and Gertman, I., 2012. On the surface circulation of the Levantine sub-basin derived from Lagrangian drifters and satellite altimetry data. Deep Sea Research Part I, 65, 46-58.

Morel A, André J-M, Andre J-M., 1991. Pigment distribution and primary production in the western Mediterranean as derived and modeled from coastal zone color scanner observations. J Geophys Res. doi:10.1029/95JC00466

Moura, D., Mendes, I., Aníbal, J., & Gomes, A. (2017). Guadiana River Estuary. Investigating the past, present and future. University of Algarve. Centre for Marine and Environmental Research (CIMA).

Mourre, B., De Mey, P., Menard, Y., Lyard, F. and Le Provost, C., 2006. Relative performance of future altimeter systems and tide gauges in constraining a model of North Sea high-frequency barotropic dynamics. Ocean Dynamics, 56, 473–486.

Moutin T, Raimbault P., 2002. Primary production, carbon export and nutrients availability in western and eastern Mediterranean Sea in early summer 1996 (MINOS cruise). J Mar Syst 33 34:273–288. doi:10.1016/S0924-7963(02)00062-3

Murphy, A. H., 1988. Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. Monthly Weather Review, 116, 12, 2417–2424.

Nerger, L., Danilov, S., Hiller, W., Schroter, J., 2006. Using sea-level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter. Ocean Dynam 56, 634-649.

Neves, R. (2007). NUMERICAL MODELS AS DECISION SUPPORT TOOLS IN COASTAL AREAS. In I. E. Gonenc, V. G. Koutitonsky, B. Rashleigh, R. B. Ambrose, & J. P. Wolflin (Eds.), Assessment of the Fate and Effects of Toxic Agents on Water Resources (pp. 171–195). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5528-7_8

Nittis, K., L. Perivoliotis, G. Korres, C. Tziavos & I. Thanos, 2006. Operational monitoring and forecasting for marine environmental applications in the Aegean Sea. Environ- mental Modelling and Software 21: 243–257

Nunes, L. J. R., Raposo, M. A. M., & Gomes, C. J. P. (2020). The Impact of Tourism Activity on Coastal Biodiversity: A Case Study at Praia da Cova Redonda (Algarve—Portugal). Environments, 7(10), 88. https://doi.org/10.3390/environments7100088

Oey, L.Y., Mellor, G.L. and Hires, R.I., 1985. A three dimensional simulation of the hudson raritan estuary. Part I-Description of the Model and Simulatio. Journal of Physical Oceanography, 15, 1676–1692.

Oke, P. R. and Schiller, A., 2007. Impact of Argo, SST, and altimeter data on an eddyresolving ocean analysis. Geophysical Research Letters, 34, L19601.

Palmer, Stephanie, Laurent Barillé, Pierre Gernez, Stephano Ciavatta, Hayley Evers-King, Susan Kay, Andrey Kurekin, et al. 2019. Earth Observation and Model-Derived Aquaculture Indicators Report. https://doi.org/10.5281/zenodo.3581506.

Papadopoulos, A., Katsafados, P., Kallos, G., & Nickovic, S. (2002). The Weather Forecasting System for Poseidon—An Overview. Journal of Atmospheric & Ocean Science, 8(2–3), 219–237. https://doi.org/10.1080/1023673029000003543

Pinardi, N., and Masetti, E., 2000. Variability of the large scale general circulation of the Mediterranean Sea from observations and modelling: A review, Palaeogeography, Palaeoclimatology and Palaeoecology, 158, 153-173.

Pisano, A., Buongiorno Nardellia, B., Tronconia, C. and Santoleria, R., 2016. The new Mediterranean optimally interpolated pathfinder AVHRR SST Dataset (1982–2012). Remote Sensing of Environment, 176, 107–116.

Pollani, A., Triantafyllou, G., Petihakis, G., Konstantinos, N., Dounias, K. and Koutitas, C., 2001. The Poseidon operational tool for the prediction of floating pollutant transport. Marine Pollution Bullettin, 43, 270–278.

Quattrocchi, G., De Mey, P., Ayoub, N., Vervatis, V.C., Testut, V., Reffray, V., Chanut, V. and Drillet V., 2014. Characterisation of errors of a regional model of the Bay of Biscay in response to wind uncertainties: a first step toward a data assimilation system suitable for coastal sea domains. Journal of Operational Oceanography, 7,2, 25-34.

Relvas, P., & Barton, E. D. (2005). A separated jet and coastal counterflow during upwelling relaxation off Cape São Vicente (Iberian Peninsula). Continental Shelf Research, 25(1), 29–49. https://doi.org/10.1016/j.csr.2004.09.006

Relvas, P., Barton, E. D., Dubert, J., Oliveira, P. B., Peliz, Á., da Silva, J. C. B., & Santos, A. M. P. (2007). Physical oceanography of the western Iberia ecosystem: Latest views and challenges. Progress in Oceanography, 74(2–3), 149–173. https://doi.org/10.1016/j.pocean.2007.04.021

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 Sci. 10, 731-744.

Robinson, A.R., Leslie, W.G., Theocharis A. and Lascaratos, A., 2001. Mediterranean Sea Circulation. In: Steele, J. H. (Eds.), Encyclopedia of Ocean Sciences. Academic Press, Oxford.

Santo, F. E., Pires, V. C., Silva, Á., & Moita, S. (2005). Caracterização Climática 2004. Instituto de Meteorologia. Ministério da Ciência, Tecnologia e Ensino Superior.

Santos, J. F., Pulido-Calvo, I., & Portela, M. M. (2010). Spatial and temporal variability of droughts in Portugal: VARIABILITY OF DROUGHTS IN PORTUGAL. Water Resources Research, 46(3). https://doi.org/10.1029/2009WR008071

Smagorinsky, J. 1963. General Circulation Experiments With the Primitive Equations. Monthly Weather Review, 91, 99–164.

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

Teles-Machado, A., Peliz, Á., Dubert, J., & Sánchez, R. F. (2007). On the onset of the Gulf of Cadiz Coastal Countercurrent. Geophysical Research Letters, 34(12), L12601. https://doi.org/10.1029/2007GL030091

Theocharis, A., Georgopoulos, D., Lascaratos, A. and Nittis, K., 1993. Water masses and circulation in the central region of the Eastern Mediterranean: Eastern Ionian, South Aegean, and Northwest Levantine, 1986–1987. Deep Sea Research Part II, 40, 1121–1142.

Theoharis, A., Balopoulos, E., Kioroglou, Kontoyiannis, H. and Iona, A., 1999. A synthesis of the circulation and hydrography of the South Aegean Sea and the Straits of the Cretan Arc (March 1994–January 1995). Progress in Oceanography, 44, 469-509.

Tsiaras, K.P., Hoteit, I., Kalaroni, S., Petihakis, G. and Triantafyllou, G., 2017. A hybrid ensemble-OI Kalman filter for efficient data assimilation into a 3-D biogeochemical model of the Mediterranean. Ocean Dynamics, 67, 673–690.

Turner, M.R.J., Walker, J.P. and Oke, P.R., 2008. Ensemble member generation for sequential data assimilation. Remote Sensing of Environment, 112, 4, 1421-1433.

Yoon, J.-H., Kawano, S., Igawa, S., 2010. Modelling of marine litter drift and beaching in the Japan Sea. Mar. Pollut. Bull. 60, 448–463.

Zavatarelli, M. and Pinardi, N., 2003. The Adriatic Sea modelling system: a nested approach. Annales Geophysicae, 21, 345–364.