Assessment of model skill in the Mediterranean Sea

by the group at **OGS** (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Italy

Contents

Summary

Regional hindcast simulations for the period 1990-2009 have been made using the OPEC regional model systems for both the biogeochemical and HTL model components. This document records the assessment of the model system's performance. The hindcast skill is evaluated using the benchmark tool developed in the framework of OPEC.

The results indicate that all regional model systems demonstrate a range of skill, depending on the variables chosen. Physical variables (e.g. T, S) are generally have the most skill) followed by chemical variables (e.g. O_2, Nutrients) then plankton variables (e.g. chlorophyll) for the coupled hydrodynamic LTL models. The HTL models have more skill for small pelagic fish (e.g. Sprat in the Baltic and Anchovy in the Aegean) than the plankton model that drives them. However the skill for larger pelagic and demersal fish is generally poor.

Introduction

A primary objective of OPEC was to set-up the ecological model system for the next generation GMES marine ecological service in European Seas. Each regional model system comprises a core coupled hydrodynamic-plankton model, a HTL component, a representation of the carbon chemistry and a data assimilation system. These have been used to perform 20yr hindcast of each region and to benchmark model performance. This document focuses on assessing the performance of the hindcast model systems with an emphasis on evaluating the performance of key policy relevant metrics. The goal is to benchmark the quality of the regional hindcasts, firstly to inform users of the data products about the skill and hence usefulness of the simulations, secondly to provide a benchmark against which future model development can be assessed.

Model skill scores are colour coded to give an indication of model skill as follows; Good = green, Moderate = Black, Poor = Red; correlation r Good > 0.75, poor < 0.20, M good = > 0.75, Poor = < 0, Reliability index Good 0.8-1.2, Poor > 2 or –ve. PBias, Good = < 20%, Poor = > 100%.

1. Description

The OGS Mediterranean model system consists of the following components:

- LTL: transport-biogeochemical model (OPATM-BFM) which is coupled with a carbonate system module; the spatial resolution of Mediteranean Sea domain is 1/8° x 1/8° in horizontal and 72 vertical levels;
- DA: 3D-variational scheme with assimilation of SeaWiFS and MODIS weekly averaged chlorophyll maps;
- HTL: Ecopath with Ecosim for the Adriatic Sea sub-region.

Atmospheric and physical forcing for the hindcast simulation are derived from MyOcean Mediterranean Forecast System (MFS) that is based on NEMO model managed by INGV (Italy). The coupling of LTL and HTL models is provided with an off-line integration in agreement with the scheme presented in Libralato and Solidoro (2009). This document presents the results of the hindcast simulation and the validation of the benchmark variables. Physical reanalyses released by MyOcean (version V3) cover the period 1999-2013, and were used for the hindcast simulation. The hindcast run of the period 1990-1998 was performed using a validated physical forcing data set which differs from that one derived from MyOcean products. This was due to a delay in the publication of the Mediterranean reanalyses product for the 1990s by the MyOcean partner (version V4). Since physical forcings of the two periods show significant differences in some important Mediterranean Sea circulation features, and since the reanalyses and the following project phase (REA experiments) will be based on MyOcean products, we have specifically focused on MyOcean products, using hindcast results of 1990-1998 as a spin-up, while they were not used for computing the statistics of OPEC target variables.

The hindcast simulation was carried out at the CINECA supercomputer facility (Italy) using 2048 cores on the supercomputer IBM BG/Q "FERMI" (performance: 10 years simulated in around 27 hours).

2. Data for validation

Reference data for surface chlorophyll consist of 1999-2010 SeaWiFS satellite observations made available through the MyOcean consortium (GOS-ISAC-CNR via the Ocean Colour Thematic Assembly Centre).

In situ data for dissolved inorganic nutrients (nitrate, phosphate, silicate), oxygen and chlorophyll-a consist of EU/MEDAR/MEDATLAS II for the period 1990-2000 (Manca et al., 2004), and of NODC-OGS databases for 2000-2011. Table 1 reports the list of oceanographic cruises and datasets gathered within OPEC task 2.3. The spatial distribution of data is shown in Figure 1.

Carbonate system variables data are available only for few campaigns, namely Meteor51 cruise (year 2001), Boum2008 (year 2008) and Sesame cruises (year 2008).

For HTL variables, data regard midyear biomass in metric tonnes (t) for the area under consideration (the Adriatic Sea -GFCM GSA 17- from Trieste to Vieste latitude), and are available for sardine (*Sardina pilchardus*) from 1975 to 2006, and for anchovy (*Engraulis encrasicolus*) from 1976 to 2006 (Santojanni et al., 2003, 2005, 2006a, 2006b, 2006c). Catch data (t) per each year from 1975 to 2006 for sardine and from 1976 to 2006 for anchovy are also available.

Table 1 List of data sets and oceanographic cruises used for the validation of model results.

Dataset name	**Period**	**Area**
SINAPSI 3,4	2002-2003	East Med.
JGOFS-FRANCE	1999	West Med.
BIOPT 6	2006	East Med.
DYFAMED	1998-2007	NorthWestMed
RHOFI 3,2,1	2001-2003	Ligurian Sea
NORBAL 1, 2, 3, 4	2000-2003	Algero Sea
CIESM SP1,SP2,SP3	1998-2006	Mediterranean
MELISSA	2004, 2007	Western Med.

MEDGOOS 2, 3, 4, 5	2001-2002	Mediterranean
METEOR 51	2001	Western Med.
METEOR 84/3	2011	Mediterranean

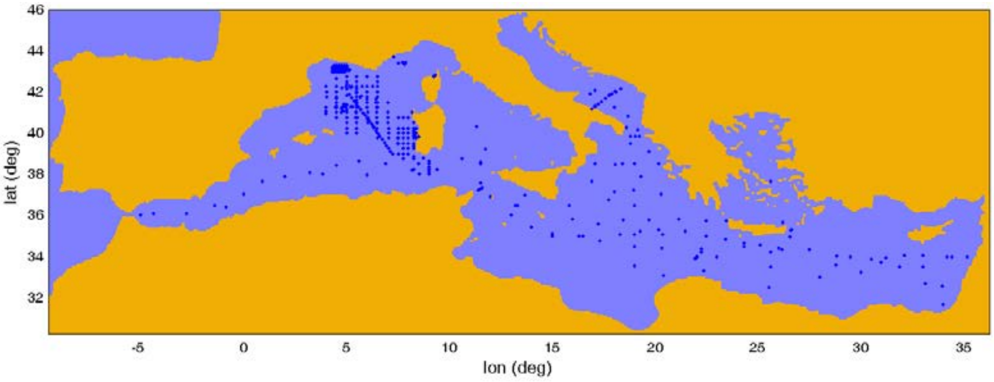


Figure 1. Map of gathered data in the NODC-OGS data set for period 2000-2011.

2.1 Comparison of surface chlorophyll against satellite data

Model and satellite observation mean trends for a selection of Mediterranean sub-regions (defined in Figure 2) are shown in Figure 3. The results of the benchmarking tool over the Mediterranean Sea (where the match-ups have been evaluated on surface chlorophyll concentration data corresponding to grid points characterized by depth beyond 200m) are shown in Figure 4.

Figure 3 shows that modelled mean seasonal cycles are consistent with observations. Model tends to slightly underestimate reference data in summer in all sub-regions shown here. Simulated values are on average 0.05 mg chl m^{-3} lower than observations. In Eastern sub-regions winter simulated values are 0.05-0.1 mg chl m^{-3} higher than satellite data.

On average, considering the benchmarking evaluated on the whole Mediterranean Sea and along the period 1999-2010 (more than 1.6 million of match-ups), the Taylor diagram (Figure 7) shows that correlation coefficient is about 0.6 and unbiased RMSE is 0.083 mg chl m^{-3}. The standard deviation of the model (0.097 mg chl m^{-3}) is slightly larger than that of the reference (0.086 mg chl m^{-3}). The target diagram (Figure 7) shows that normalized bias is very small and that RMSE has same magnitude than the reference standard deviation, meaning that, on average, the difference between model and reference is similar to the reference data variability.

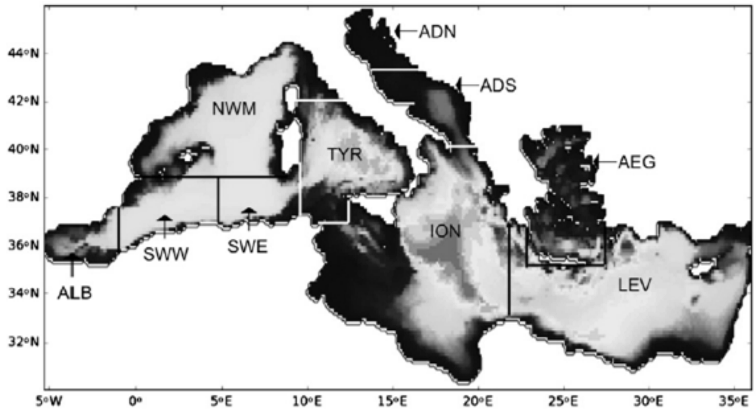


Figure 2. Model domain of the Mediterranean Sea with selected sub-regions of investigation: ALB = Alboran Sea, SWM = south-western part of western Mediterranean Sea, SWE = south-eastern part of western Mediterranean Sea, NWM = north-western Mediterranean Sea, TYR = Tyrrhenian Sea, ADN = northern Adriatic Sea, ADS = southern Adriatic Sea, AEG = Aegean Sea, ION = Ionian Sea, LEV = Levantine basin.

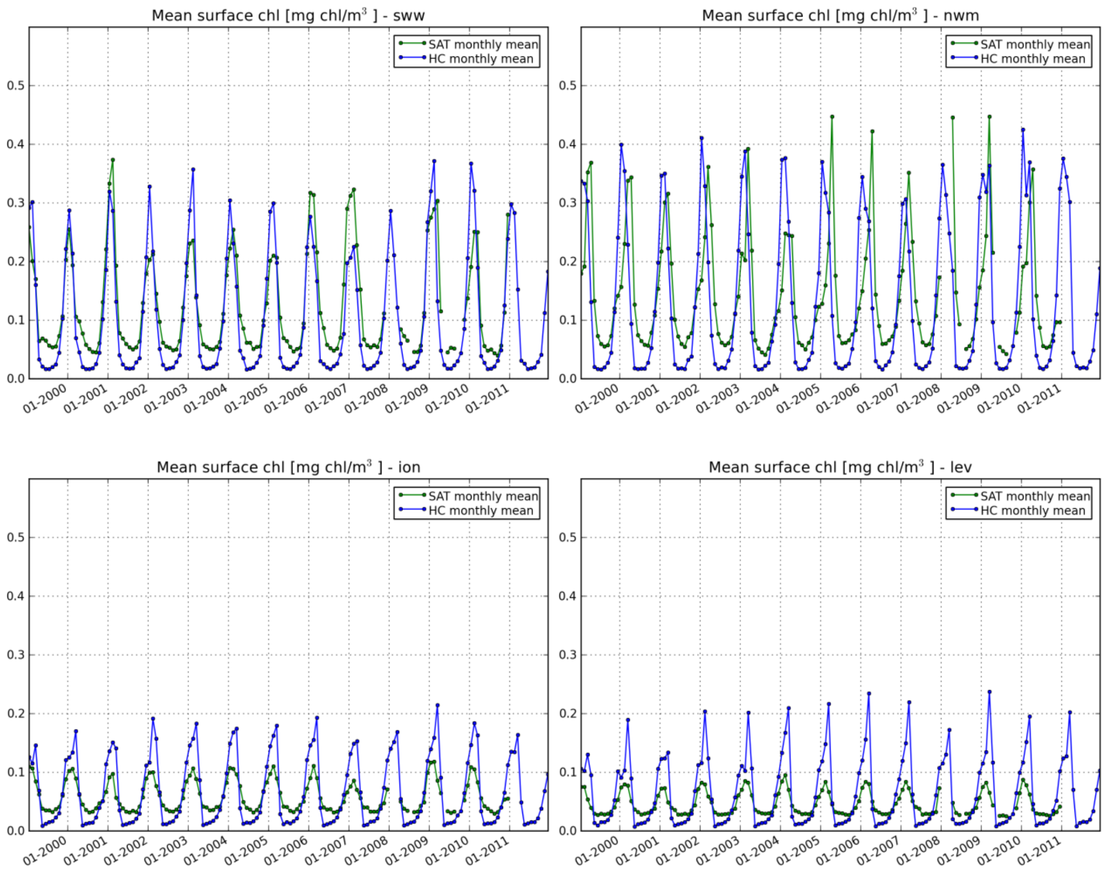


Figure 3. Trends of mean surface chlorophyll of hindcast simulation (blue line) and satellite data (green line) in selected sub-regions of the Mediterranean Sea.

2.2 Comparison of simulated chlorophyll, nutrients and oxygen against in-situ data

The figures in the present section report the evolution of the mean seasonal vertical profiles in the period 1999-2011 (Figure 4-6, grey-scale lines) for selected Mediterranean sub-regions.

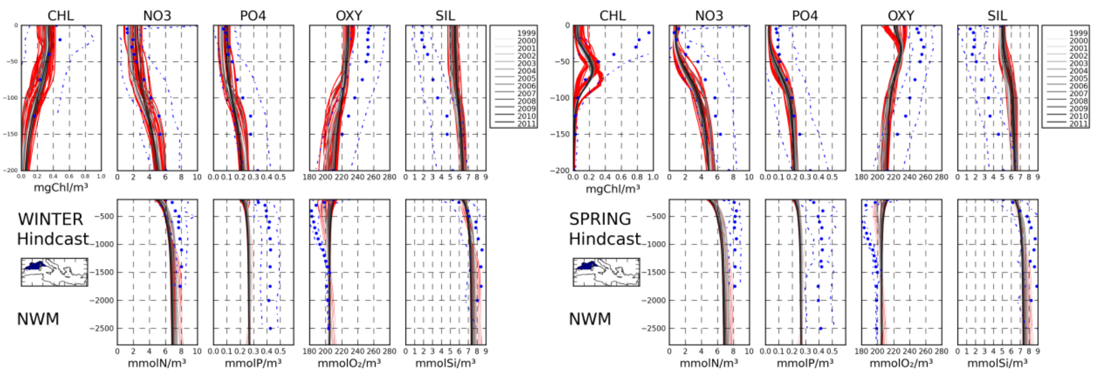
The mean vertical profile is computed averaging all grid points with a water column deeper than 200 m. The dispersion of the monthly profiles (Figure 4-6, red lines) gives information on the temporal variability. Mean climatological profiles computed on 2000-2010 data are given by blue dots (dashed lines correspond to standard deviations). This comparison allows, even if qualitatively, to appreciate the model's capability to reproduce the main biogeochemical features of the Mediterranean Sea.

Simulated deep chlorophyll maximum is at 70-80 meters during spring and summer with values of 0.3 mg chl m^{-3} in western sub-regions, whereas it deepens at 100 meters in the eastern sub-regions. During winter, surface chlorophyll concentrations are 0.2-0.3 mg chl m^{-3} in western sub-regions and 0.1-0.2 mg chl m^{-3} in eastern sub-regions. Negligible surface values are simulated in the other seasons. Simulated mean seasonal features are qualitatively consistent with the climatological values (with the notable exception of spring in NWM, where model is not able to reproduce the larger surface layer concentrations, possibly related to the specific reference data set which included coastal observations).

Consistently with climatological profiles, the simulated concentration of nutrients below the photic layer is almost constant throughout the period covered by the hindcast simulation with values of around 4-6 mmol m^{-3}, 0.15-0.25 mmol m^{-3} and 7-8 mmol m^{-3} for NO_3, PO_4, SiO_2, respectively. In the surface layer concentrations are close to zero for all nutrients but Silicate.

Simulated silicate concentrations are slightly higher than those of climatological profiles for the upper layer in all the sub-regions.

Simulated oxygen values for the deep layer are around 200 mmol m^{-3} consistently with climatological values. In the surface layer oxygen oscillate between 180 (summer) and 220 (winter). During summer, the maximum values along the simulated oxygen profiles are positioned just above the deep chlorophyll maximum coherently with the observations; however simulated values are slightly lower than those reported by climatological profiles.



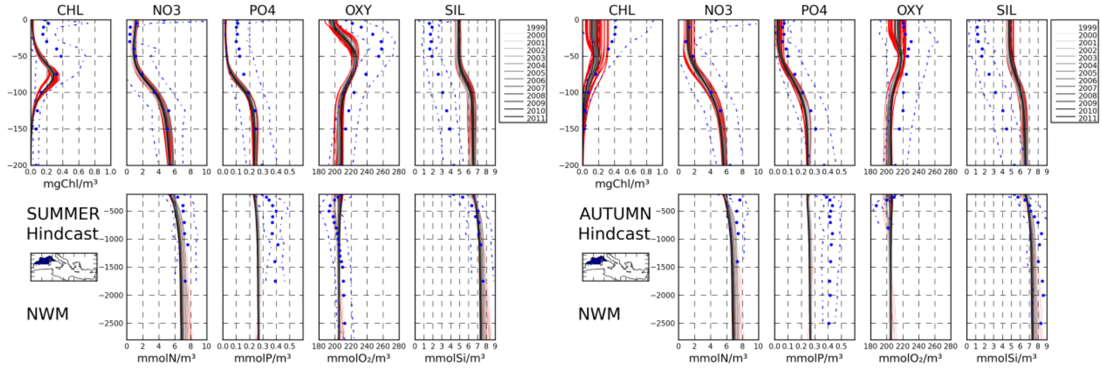


Figure 4. Mean seasonal (gray-scale lines) and monthly profiles (red lines) of chlorophyll, nitrate, phosphate, silicate and dissolved oxygen in the NWM sub-region. Climatological profiles of mean (blue dots) and standard deviation range (blue dashed-dotted lines) are included.

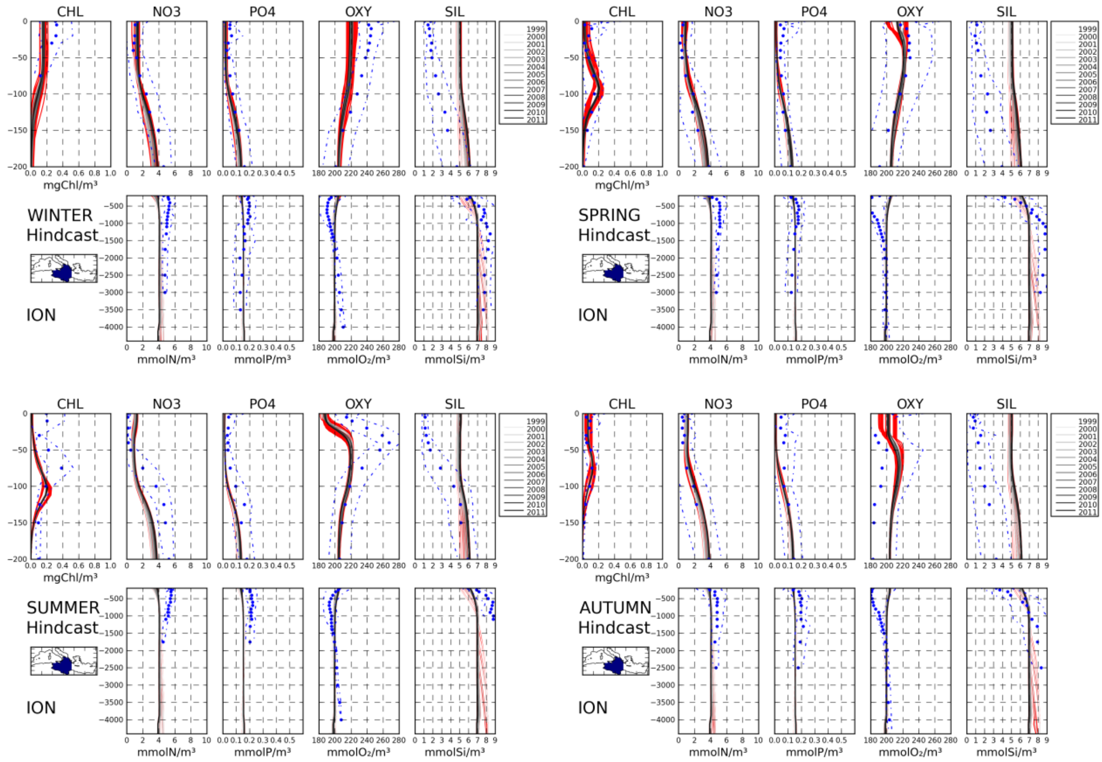


Figure 5. Mean seasonal (gray-scale lines) and monthly profiles (red lines) of chlorophyll, nitrate, phosphate, silicate and dissolved oxygen in the ION sub-region. Climatological profiles of mean (blue dots) and standard deviation range (blue dashed-dotted lines) are included.

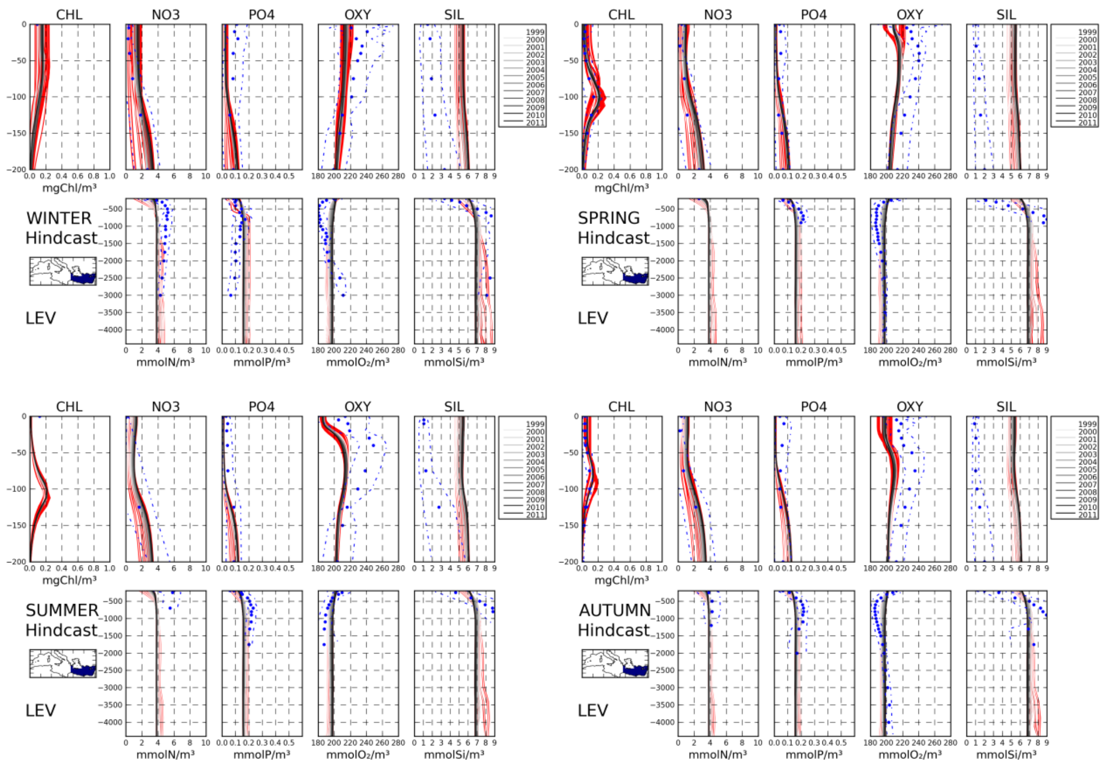


Figure 6. Mean seasonal (gray-scale lines) and monthly profiles (red lines) of chlorophyll, nitrate, phosphate, silicate and dissolved oxygen in the LEV sub-region. Climatological profiles of mean (blue dots) and standard deviation range (blue dashed-dotted lines) are included.

A quantitative assessment of the hindcast simulation has been performed using the benchmark tool for the data available in the period 2000-2010. The results of the skill assessment for the entire period are plotted in Figures 7-10 for nitrate, phosphate, silicate and dissolved oxygen. The tool has been applied considering two layers: upper layer (0-200m) and deep layer (200-bottom), in order to assess the skill of the model both within the photic zone and in the deeper part of the water column, where biological process are almost absent and physics dynamics slower.

Statistics computed by the benchmark tool are then summarized in Table 2.

Results of the benchmark tool suggest that the hindcast run has some good skill in reproducing the in-situ data for most of the variables. For surface chlorophyll, phosphorus (whole column) and nitrate and oxygen in the deeper layer the model efficiency values are positive, showing that the model-data differences are generally lower than the data variability (between 10 and 80%). For oxygen and nitrate in the upper layer and silicate the model efficiency is slightly negative. In particular, silicate in the surface layer shows the worst performance among variables. A possible overestimation of the terrestrial input and initialization is the cause of the unsatisfactory model capability of simulating silicate.

Percentage model biases are lower than 1% for surface chlorophyll and oxygen in the surface layer, and maintain between 4-5% for the other variables, while increase, in absolute value, around 40% for phosphate in the surface layer and silicate in the deeper layer, and above 100% for silicate in the surface layer. A link of the previous skill indicators with correlation is not direct for all the variables, however, the best model skill is reached with chlorophyll, and good values of correlation (higher than 0.5) are shown for oxygen in the upper layer, phosphate, and silicate in the deeper layer, with model efficiency values not larger than 0.24 in absolute value. Such heterogeneous features are strongly related to the relatively limited data set presently available, mainly provided by research campaigns localized in time and space. Density plots of Figures 7-10 show quite clearly this aspect, especially in terms of a substantial dispersion of data: model is generally unable to represent the full variability of reference data, especially in the deeper layer. Since most of reference data were measured during spring (March to May) or autumn (September to December), this aspect could be related to a partial inefficient description of the mixing processes: detailed investigations on this aspect (e.g. analysing only reference data relative to specific areas/periods) would be useful to get better insight on the model performance. Further, a possible improvement in the skill can be reached by setting upper and lower limits for reference data (often significantly related to the dynamics of measurement area/period).

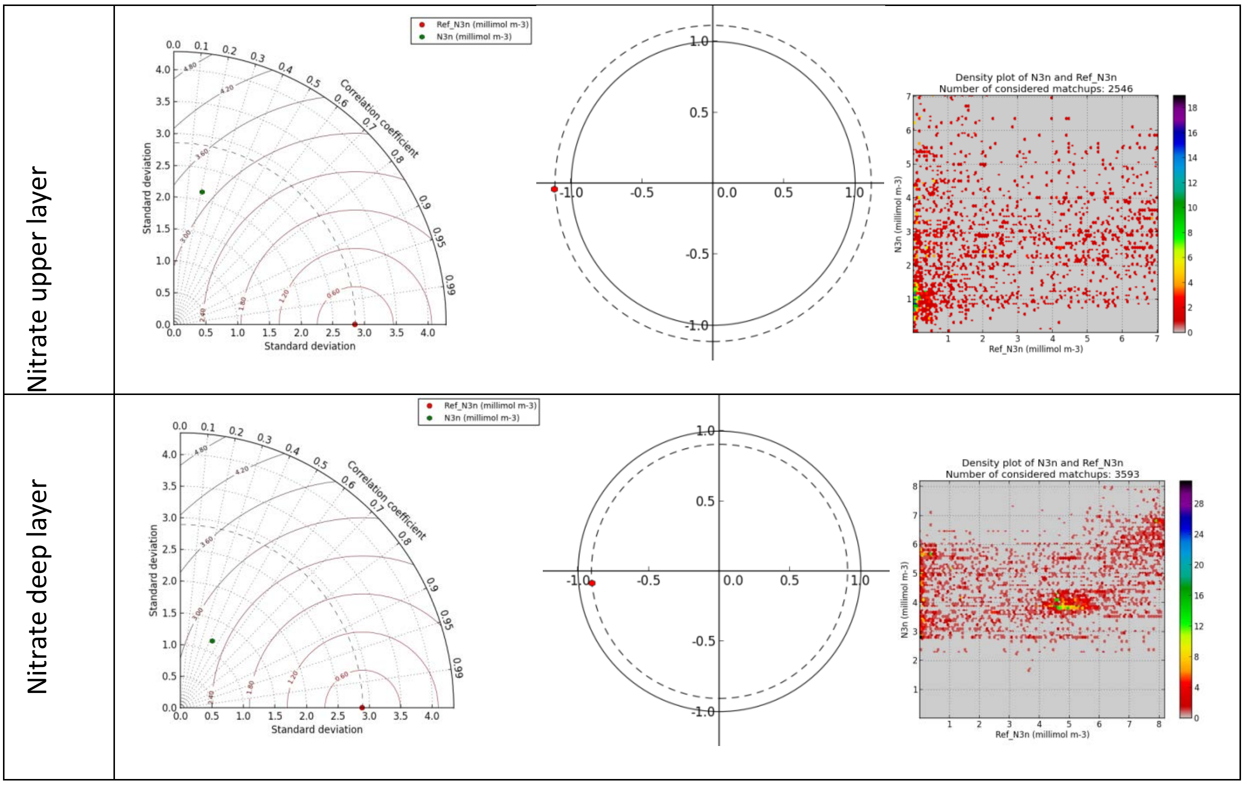


Figure 7. Results from the benchmarking tool for the whole Mediterranean Sea (period 1999-2011) for the upper layer (top row, 2546 match-ups) and for the deeper layer (bottom row, 3593 match-ups): Taylor diagram (left), target diagram (middle), density plot (right).

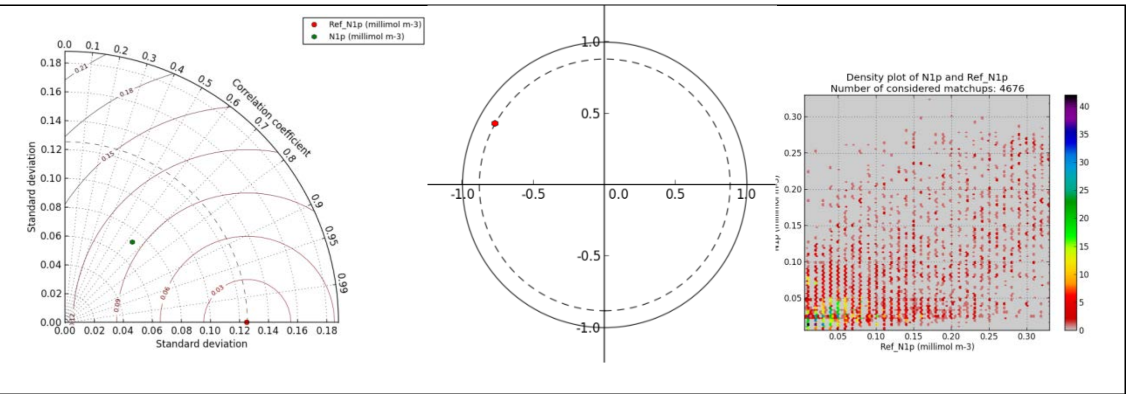
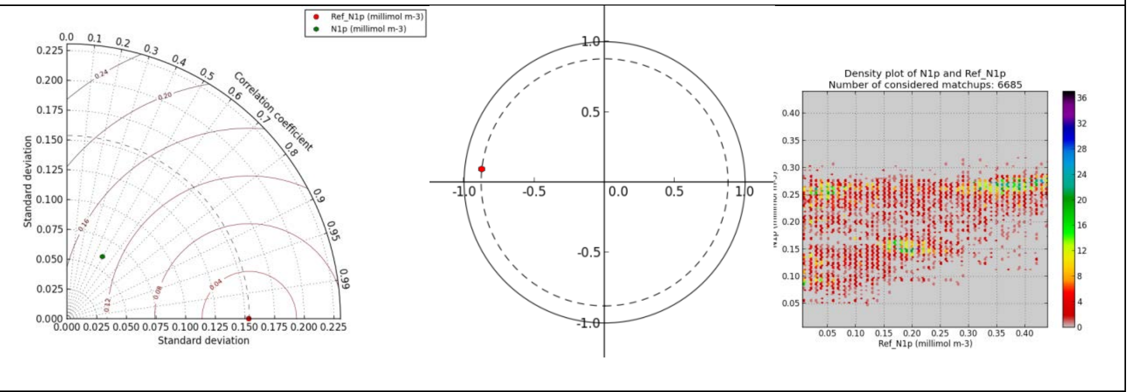
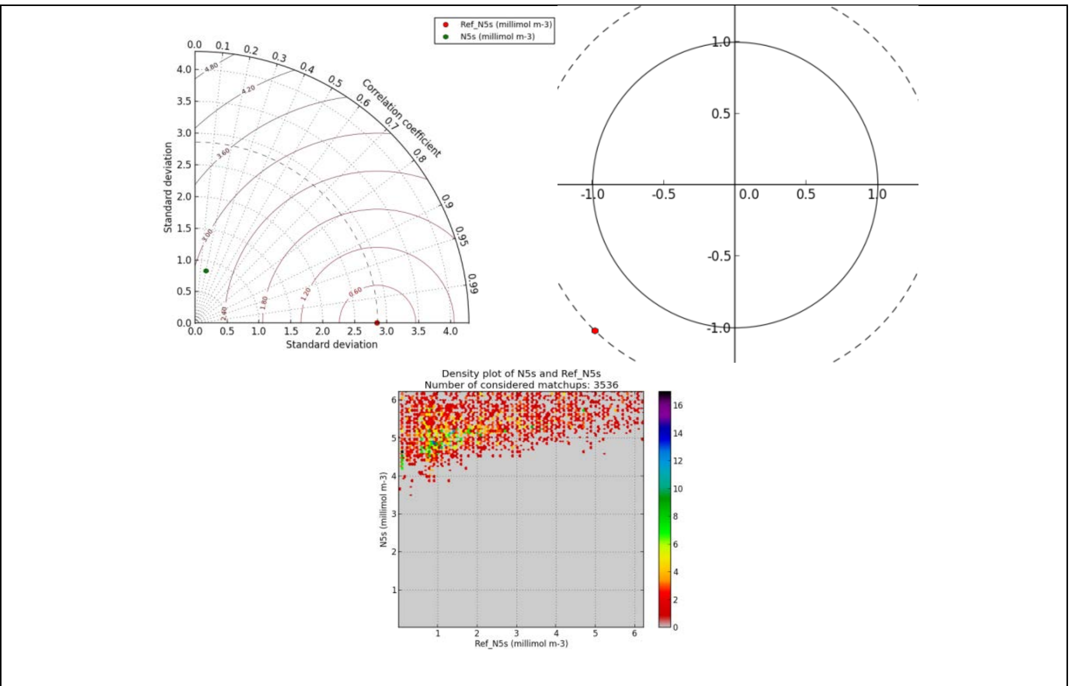
Phosphate upper layer	
Phosphate deep layer	

Figure 8. Results from the benchmarking tool of Phosphate for the whole Mediterranean Sea (period 1999-2011) for the upper layer (top row, 4676 match-ups) and for the deeper layer (bottom row, 6685 match-ups): Taylor diagram (left), target diagram (middle), density plot (right).

Silicate upper layer	

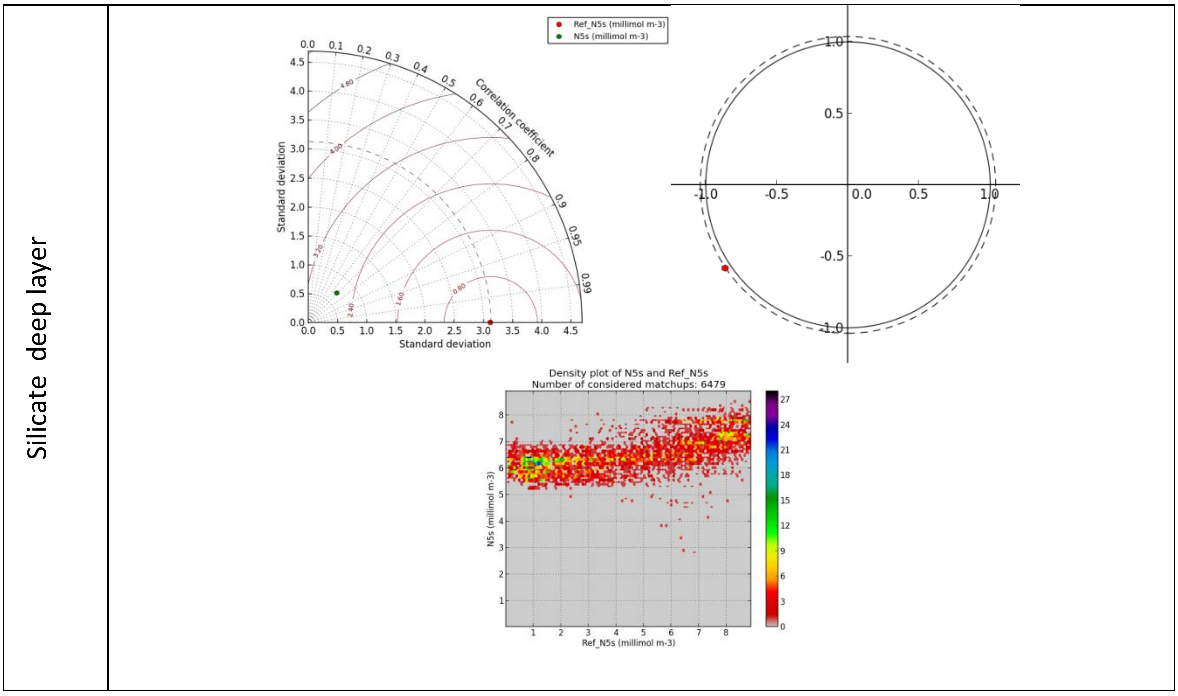


Figure 9. Results from the benchmarking tool of Silicate for the whole Mediterranean Sea (period 1999-2011) for the upper layer (top row, 3536 match-ups) and for the deeper layer (bottom row, 6479 match-ups): Taylor diagram (left), target diagram (middle), density plot (right).

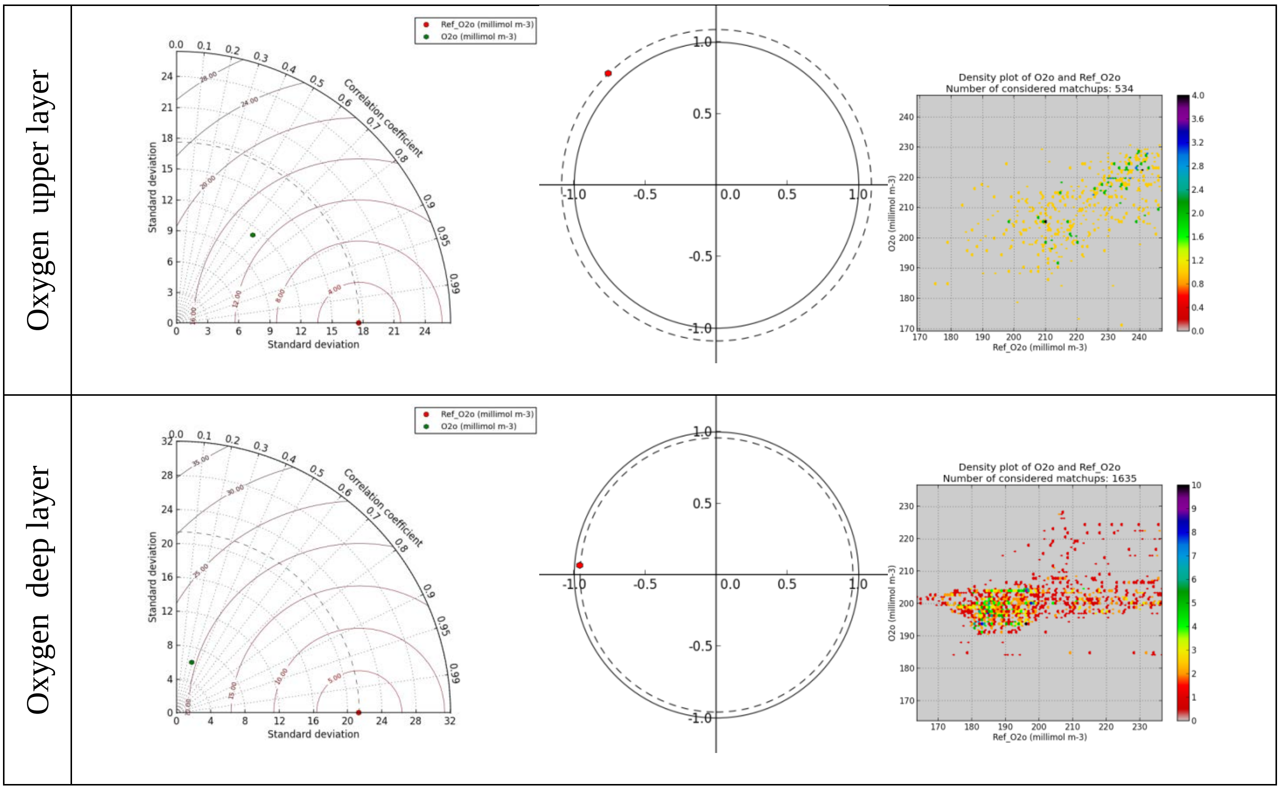


Figure 10. Results from the benchmarking tool for the whole Mediterranean Sea (period 1999-2011) for the upper layer (top row, 534 match-ups) and for the deeper layer (bottom row, 1635 match-ups): Taylor diagram (left), target diagram (middle), density plot (right).

Table 2 Skill assessment statistics for hindcast simulation. Skill assessment computation of all variables, but chlorophyll, is performed for two layers: surface layer (0-200) and deep layer (200m-bottom of water column). Skill assessment of Chlorophyll is performed for surface layer only. Model skill scores are colour coded to give an indication of model skill as follows; Good = green, Moderate = Black, Poor = Red; correlation r good > 0.75, poor < 0.20, ME good = > 0.5, Poor = < 0, Reliability index good 0.8-1.2, poor > 2 or –ve. PBias, good = < 20%, poor = > 100%.

Region: **Mediterranean (OGS)**
Model: OPATM-BFM-EwE
Met/Circulation forcing: ECMWF/MyOcean V3 Reanalysis
Hindcast time period: 1999-2010
Contact: ssalon@ogs.trieste.it (Stefano Salon), gcossarini@ogs.trieste.it (Gianpiero Cossarini), slibralato@ogs.trieste.it (Simone Libralato)

		Hindcast					
Variable	Unit	Model Efficiency	% Model Bias	Pearson Correlation Coefficient	RMSD	Unbiased RMSD	Reliability Index
CHL satellite	mgchl m^{-3}	0.08	**-0.85**	0.60	0.08	0.08	1.45
DOXY surf.	mmolO2 m^{-3}	**-0.19**	**6.09**	0.65	19.20	13.37	**1.04**
DOXY deep	mmolO2 m^{-3}	0.08	**0.68**	0.29	20.49	20.44	**1.04**
NTRA surf	mmolN m^{-3}	**-0.24**	**-4.27**	**0.21**	3.18	3.18	2.13

NTRA deep	mmolN m^{-3}	0.18	-5.77	0.43	2.62	2.61	1.98
PHOS surf	mmolP m^{-3}	0.22	39.78	0.64	0.11	0.10	1.53
PHOS deep	mmolP m^{-3}	0.23	6.00	0.50	0.14	0.13	1.52
SLCA surf	mmolSi m^{-3}	-1.01	-115.35	0.21	4.05	2.81	1.94
SLCA deep	mmolSi m^{-3}	-0.08	-38.36	0.69	3.25	2.69	1.66

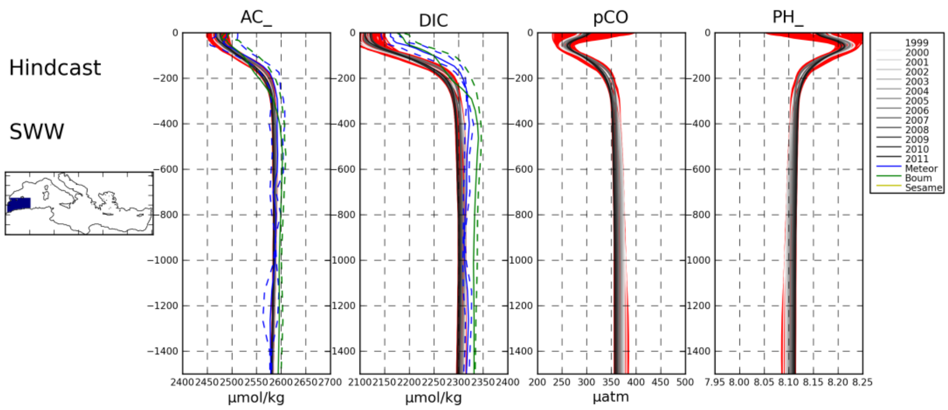
2.3 Validation of Carbonate system

Mean annual profiles of the carbonate system are shown in Figure 11 for selected sub-regions. Mean profiles of the cruises are reported in the plots. Since the scarcity of the available data the benchmarking tool has not been applied.

The most important feature of carbonate system in the Mediterranean is the West to East gradients of Alkalinity and DIC at surface.

The qualitative comparison with the data doesn't falsify the model capability to reproduce this feature.

The inflow of low alkalinity and DIC water from Gibraltar and the input of alkalinity and DIC from the eastern rivers and from the Black Sea are the main drivers of the this permanent characteristic. Consequently the pH is slightly lower in the Eastern sub-regions than in the Western sub-regions. Further since the longitudinal gradient is overlapped by a North-to-South temperature gradient, the pCO_2 values in the eastern gradient is slightly higher than that of western subregions, making the western sub-regions a more efficient sink of atmospheric CO_2.



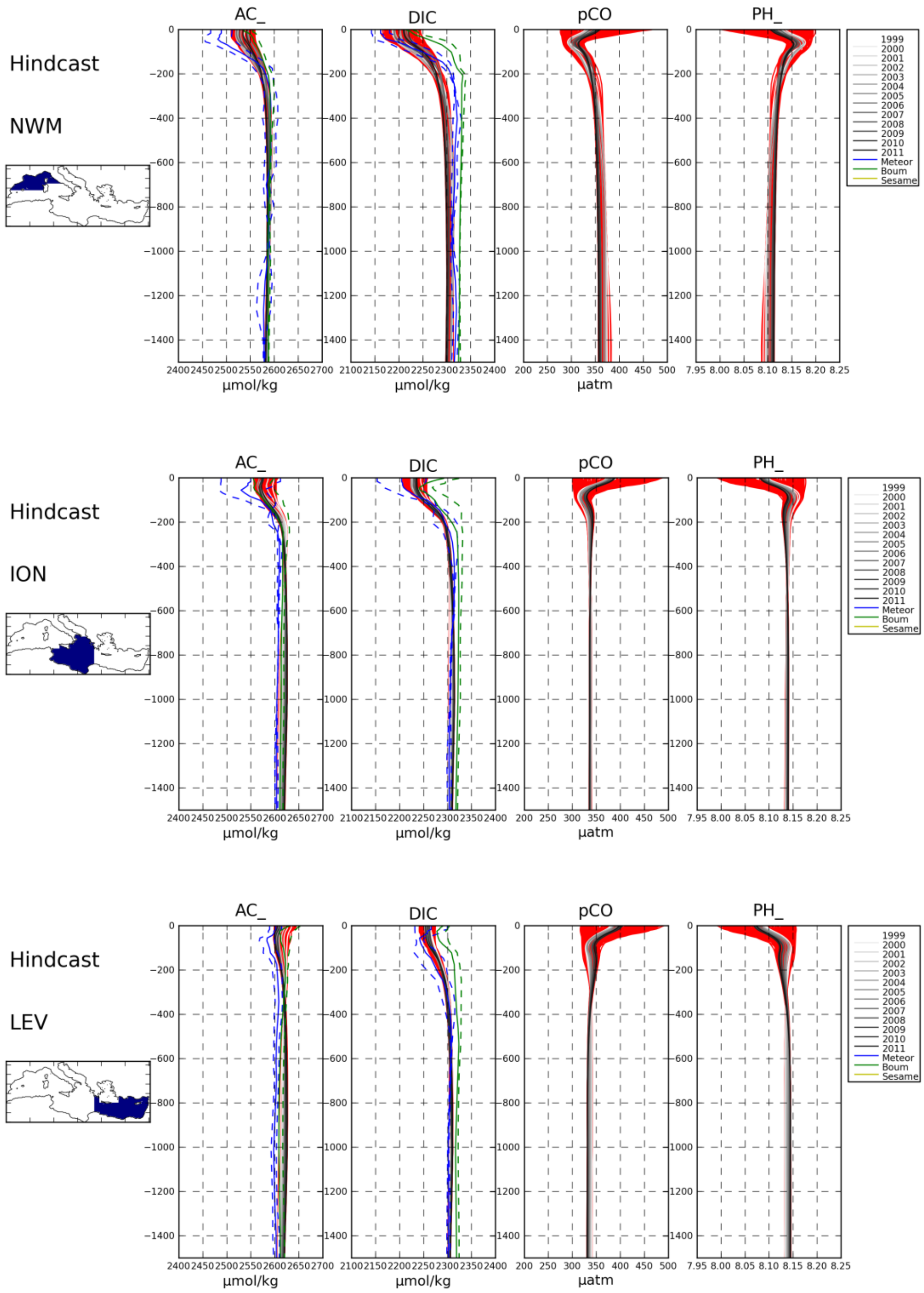


Figure 11. Mean annual profiles (grey lines) of Alkalinity (AC_), dissolved inorganic carbon (DIC) pCO2 (pCO), pH (pH_) for selected Mediterranean sub-regions. Monthly profiles (red lines), mean and standard profiles for available datasets (green and blue lines).

2.4 Summary of hindcast experiment and lessons learnt

Results of the benchmark simulation evidence that the Mediterranean Model System has a good capability of reproducing spatial gradients, seasonal cycle and interannual variability for most of the variables, even if some systematic errors are present. For surface chlorophyll, phosphorus (whole column) and nitrate and oxygen in the deeper layer the model-data differences are generally lower than the data variability (between 10 and 80%). Simulations of the upper layer dynamic for oxygen and nitrate are slightly less good. Simulation of silicate concentration in the surface layer shows the worst performance among variables. A possible overestimation of the terrestrial input and initialization is the cause of the unsatisfactory model capability of simulating silicate.

2.5 Validation of the HTL model

2.5.1 Description of the HTL Model System

The regional model system comprises the BFM-OPATM LTL model for the Mediterranean Sea and the Ecopath with Ecosim (EwE) model for the Adriatic Sea as a focal region.

The EwE model was run between 1989 and 2010. The model consisted in 47 functional groups developed on the basis of published works (Coll et al., 2007; 2009) updated for initial conditions (biomasses, catches, discards) and forcing functions, and built in nutrient units (mgP/m^2). Moreover, the initial model was extended to represent LTL compartments by adding 2 groups of plankton producers (phytoplankton and picoplankton), two groups of zooplankton (microzooplankton and mesozooplankton), bacterioplankton, phospate, dissolved and particulate organic phosporous (DOP and POP), and input of nutrient (phosphate). These functional groups in the extended EwE have been set up using OPATM-BFM hindcast run to define initial conditions and average exchange flows (grazing, natural mortality, excretion, sinking). In the extension of the EwE model to represent the groups from OPATM-BFM an adjustment of diet of planktivores was necessary: their predation on plankton producers was split according to detailed information whenever possible and to total annual average production otherwise. The same was applied to predators of zooplankton groups.

The hindcast run of the extended EwE model for the years 1989-2010 was driven using:

- fishing effort expressed as horse power changes for purse seiners, otter and bottom trawlers, midwater trawlers and tuna fleets; data were taken from published sources (Sistema Statistico Nazionale, IREPA onlus);
- fishing mortality for anchovy and sardine as estimated from Virtual Population analysis (SAC-GFCM);
- productivity changes for phytoplankton, picoplankton and bacterioplankton as derived from OPATM-BFM hindcast run (including the spin-up period).

A check of the capability of the extended model to represent biogeochemitry processes was done.

An anomaly function representing an estimation of input/output of phosphate in the system was estimated by the extended EwE model by fitting phospate and phytoplankton outputs from extended EwE with OPATM-BFM results. This adjustment, according to Libralato and Solidoro (2009) is necessary to account the different forms of aggregation necessary when performing the extension of the model. Reparametrization, aggregation of functional groups, aggregation in space and in time, are therefore corrected through the anomaly function. The resulting hindcast End-to-End simulation was overall performing satisfactorily to represent the LTL groups and to represent the HTL dynamics. Some of the outputs of the hindcast simulation (1989-2010) of the extended EwE model are exemplified in Figure 12, where the dynamics of some groups are represented.

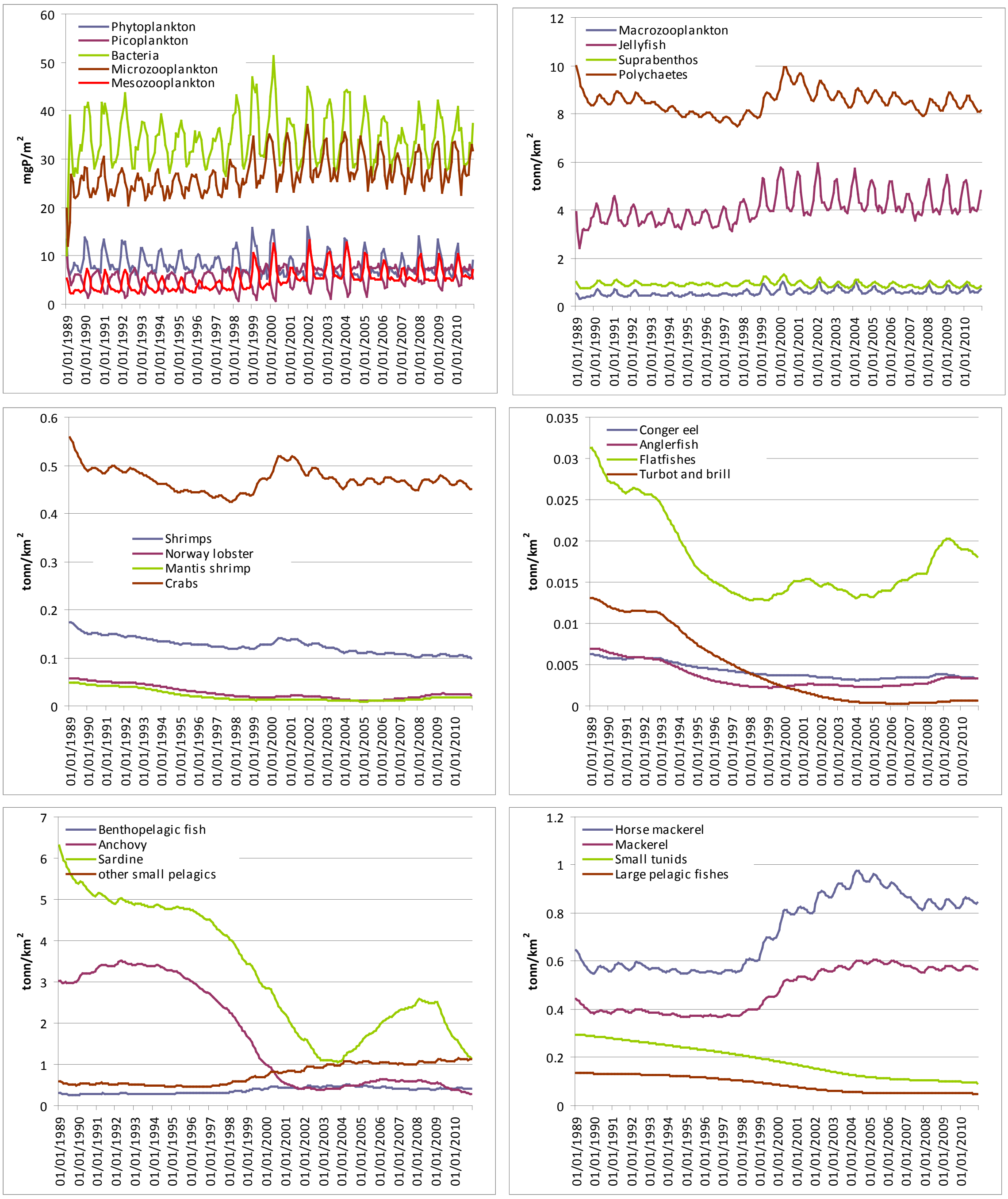


Figure 12. Temporal trends of some of the functional groups of the extended food web model: a) plankton groups; b) invertebrates; c) commercial invertebrates; d) demersal predatory fishes; e) small pelagics; f) larger pelagic fish.

The outputs of the hindcast simulation (1989-2010) of the extended EwE model was confronted with available data for sardine and anchovy estimated using stock assessment approaches (virtual population analysis reconstructed from SAC-GFCM). The simulated total biomass (tonn km^{-2}) lies within the range of observation for both sardine (0.5-6 tonn km^{-2}) and anchovy (1-3.5 tonn km^{-2}) , but the model represents the main direction of changes occurred in the past two decades with a delay in the order of 2-3 years for anchovy and 3-4 years for sardine. Overall the model skill for these two variables is poor.

2.5.2 Results

Table 3: summary table of model skill for core metrics.

Notes - * - Bias is zero because means of observed are scaled to means of data. + - detrended data.

Region**: Mediterranean (OGS)** Model: Ecosim integrated with OPATM-BFM Met forcing: Hindcast time period: 1989-2010 Contact: slibralato@ogs.trieste.it							
		Hindcast					
Variable	**Unit**	**Model Efficiency**	**% Model Bias**	**Pearson Correlation Coefficient**	**Spearman Correlation Coefficient**	**Unbiased RMSE**	**Reliability Index**
anchovy	tonn km^{-2}	-2.2746	3.2917	0.1743	0.1541	1.3535	1.518
sardine	tonn km^{-2}	-0.5568	-1.0841	0.1743	0.1337	2.3847	1.6387

2.5.3 Summary and Lessons Learnt

The results (Table 3 and Fig. 13) show low correlations for the two datasets. This is to be expected as only some of the effects of lower trophic levels on the HTL Mediterranean model are included. For instance effect of LTL on recruitment is represented implicitly through influences of LTL biomass on the productivity of their predators. In future model representation it would be necessary to include effects of timing of larval bloom on spawning and hence future recruitment, possibly including direct effects of temperature on feeding and survival of HTL species independent of effects via LTL species.

It is important to note that the Adriatic system went through important changes in the last decades in many of its ecosystem components, including the anchovy and sardine biomass. This is possibly due to fishing impacts but also to environmental changes. Therefore, it is quite difficult to fully represent the large ecosystem changes occurred. The time response of the model was a bit under responsive for both anchovy and sardine with a delay in the order of 2-3 years for anchovy and order 3-4 years for sardine. In general therefore, it seems that the model is slow responsive of LTL and fishing changes.

Given the demonstrated capability of the extended model to reproduce LTL dynamics, such delay might be resulting from contribution of different important processes. First, the recruitment is a process whose changes might have high and fast impacts on the population dynamics, while implicit representation smooth down this effect. Second the model represents anchovy and sardine as a unique functional group without age classes (multi-stanza): such simplified representation might fail in representing match-mismatch of younger classes of the population, and thus the propagation of fast changes in population density. Third, other factors such as temperature and salinity, but also nutrient changes over time due to river-runoff might deeply affect changes over time of species aggregation and potential for growth. Finally the fishing effort, i.e., the main driver of the ecosystem changes over time, is necessarily affected by errors too and its uncertainty might have overall importance on all ecosystem dynamics. In fact, changes in horse power of the different fleets was used to drive the fishing mortalities to most of the species, nevertheless, although the best estimator available for fishing effort changes, this might not be accurate enough.

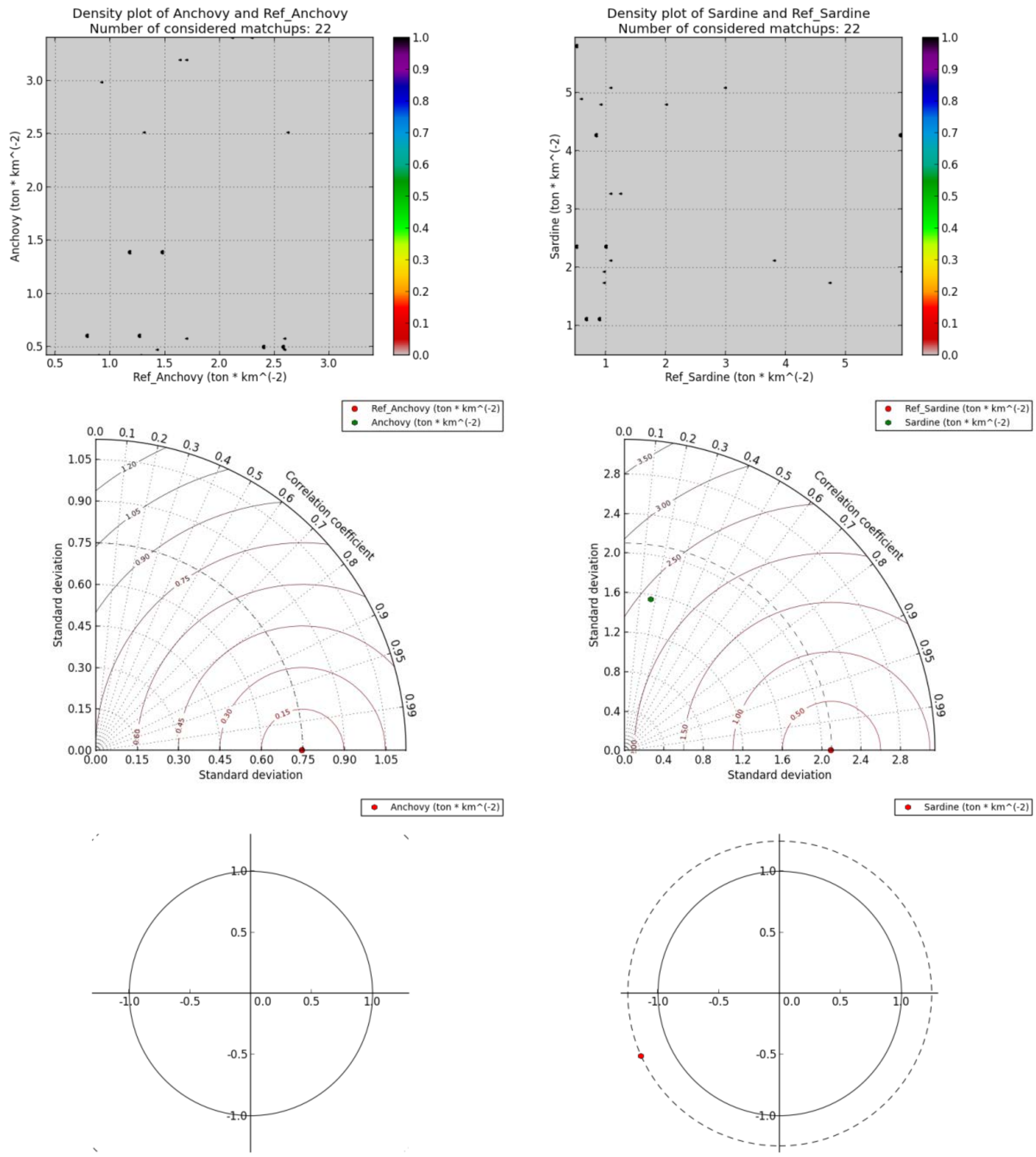


Figure 13. Benchmarking report for the HTL variables. Density plot, Taylor and Target Diagrams for anchovy (left) and sardine (right) for the Adriatic sea. Suggest colour coding green = good, yellow = ok, red = poor.

Although results show that LTL changes are propagating on the HTL groups with the E2E method developed in Libralato and Solidoro (2009), the dynamics of target species anchovy and sardine is poorly represented by the model that is showing a delay in model vs data. This is due to a series of possible reasons including the fact that HTL model structure that only implicitly represents population cohort and recruitment dynamics. In the next future, therefore, some further analyses will be done in order to improve age classes representation for anchovy and sardine possibly including some recruitment function. Moreover, a better representation of fishing effort seems necessary.