Ocean Science

Mediterranean Forecasting System: forecast and analysis assessment through skill scores

M. Tonani[1]**, N. Pinardi**[2]**, C. Fratianni**[1]**, J. Pistoia**[1]**, S. Dobricic**[3]**, S. Pensieri**[4]**, M. de Alfonso**[5]**, and K. Nittis**[6]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy
[2]University of Bologna, Corso di Scienze Ambientali, Ravenna, Italy
[3]Centro euro-Mediterraneo per i Cambiamenti Climatici, Bologna, Italy
[4]Consiglio Nazionale delle Ricerche-ISSIA, Genova, Italy
[5]Puertos del Estado, Madrid, Spain
[6]Hellenic Centre for Marine Research, Athens, Greece

Abstract. This paper describes the first evaluation of the quality of the forecast and analyses produced at the basin scale by the Mediterranean ocean Forecasting System (MFS) (http://gnoo.bo.ingv.it/mfs). The system produces short-term ocean forecasts for the following ten days. Analyses are produced weekly using a daily assimilation cycle. The analyses are compared with independent data from buoys, where available, and with the assimilated data before the data are inserted. In this work we have considered 53 ten days forecasts produced from 16 August 2005 to 15 August 2006.

The forecast skill is evaluated by means of root mean square error (rmse) differences, bias and anomaly correlations at different depths for temperature and salinity, computing differences between forecast and analysis, analysis and persistence and forecast and persistence. The Skill Score (SS) is defined as the ratio of the rmse of the difference between analysis and forecast and the rmse of the difference between analysis and persistence. The SS shows that at 5 and 30 m the forecast is always better than the persistence, but at 300 m it can be worse than persistence for the first days of the forecast. This result may be related to flow adjustments introduced by the data assimilation scheme. The monthly variability of SS shows that when the system variability is high, the values of SS are higher, therefore the forecast has higher skill than persistence.

We give evidence that the error growth in the surface layers is controlled by the atmospheric forcing inaccuracies, while at depth the forecast error can be interpreted as due to the data insertion procedure. The data, both in situ and satellite, are not homogeneously distributed in the basin; therefore, the quality of the analyses may be different in different areas of the basin.

1 Introduction

The aim of this study is to evaluate the accuracy of the forecast produced at the basin scale by the Mediterranean ocean Forecasting System (MFS), developed during MFS-Pilot Projetc,MFSPP (Pinardi et al., 2003) and MFS-Toward Environmental Predictions, MFSTEP projects. The existence of a forecasting system is important since the information delivered add some knowledge to the future event of interest. For an operational forecasting system there is therefore a need to do a validation of the products everywhere and for every single forecast day (Jolliffe and Stephenson, 2003).

The forecast could be used for scientific or operational application if there is an information of the accuracy of the predicted fields. The errors could be introduced by different components and is not straightforward to separate the contributions of each possible error source (Lermusiaux et al., 2006).The quality of the forecast is defined, following Murphy (1993), as a function of the differences between forecast and observations.

In this study we will assess the analyses against the available observations and the forecast against the analyses. The observations are divided into two categories, independent and quasi-independent: the former are data that do not enter
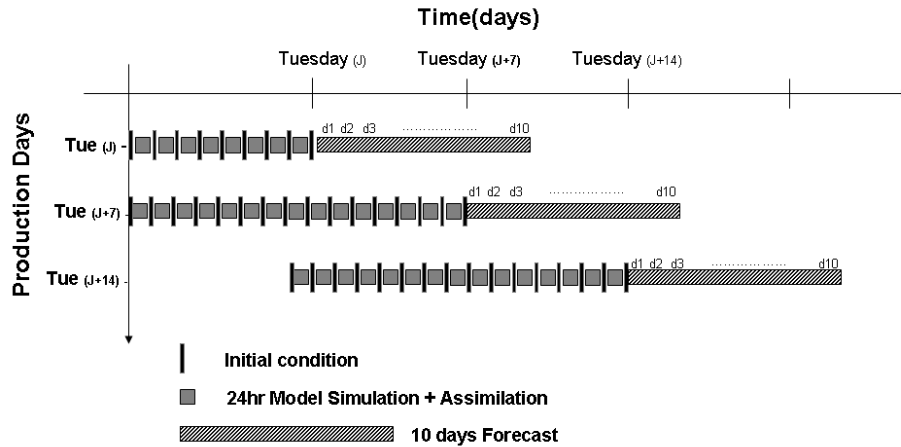
Fig. 1. MFSTEP production cycle. Every Tuesday (J) a ten-day forecast (d1, d2, d3, d4, d5, d6, d7, d8, d9, d10) is produced. It is initialised by an analysis generated by the past sequence of 15 intermittent daily data assimilation cycles.

the assimilation system while the latter are assimilated. The observations are sparse in time and space and their number is relatively low with respect to the degrees of freedom of the system. To assess the forecast quality we decided to use the analyses assuming that they represent the best estimates of the flow field, following the example of meteorology, since they contain observations and are forced with atmospheric forcing from analysis fields.

The assessment is carried out by computing the root mean square error (rmse) defined with the difference between observations and analyses or with the difference between the forecast and the analyses. In this paper we decided to use a Skill Score (SS) index following Murphy (1988) and Demirov et al. (2003), composed with the rmse, the bias and the anomaly correlations. The study concentrates on the temperature and salinity fields at selected depths (5, 30, 150, 300 and 600 m). The period considered is from August 2005 to August 2006, considering one forecast a week, i.e., 53 forecasts of ten days each. Each forecast starts on Tuesday (J) (Fig. 1) and lasts ten days from d1 to d10 starting from an analysis. Some effort has been put in the understanding of the variability of the forecast accuracy due to the seasons.

The results of this paper provide the first comprehensive assessment of the accuracy of the real time ocean forecasting system implemented in the Mediterranean Sea.

This paper is organized in the following way: Sect. 2 describes the MFSTEP forecast system and production chain. Section 3 describes the observations and the quality indices and Sect. 4 discusses the results of the analyses and forecast evaluation while Sect. 5 presents the conclusions.

2 Description of the MFSTEP forecast system and forecast production chain

The forecast production consists of the collection of in-situ and satellite data adequately pre-processed, a numerical model and the assimilation scheme. The numerical model at the finite differences and with an implicit free surface (Tonani et al., 2008) has been implemented for the Mediterranean Sea with a horizontal resolution of $1/16° \times 1/16°$ of degree and 72 unevenly spaced vertical levels. The model is forced at the air-sea interface with atmospheric fields from the European Centre for Medium-Range Weather Forecast (ECMWF) analyses and forecasts. The assimilation scheme used is a reduced order Optimal Interpolation system implemented in the Mediterranean Sea at different levels of complexity for the past ten years (Dobricic et al., 2004, 2007; Demirov et al., 2003). The assimilated data are: temperature and salinity vertical profiles from eXpandable BathyThermograph (XBT) and Argo, and Sea Level Anomalies (SLA) from altimetry. The data are collected and prepared every week (see Appendix A) on Tuesday. The first layer model temperature is relaxed toward Optimally Interpolated Sea Surface Temperature (SST, Marullo et al., 2007) derived from high resolution infrared satellite images.

The system produces analyses once a week and a ten days forecast starting from them. Figure 1 shows the daily assimilation cycle and the analyses produced for the previous 15 days (from J-15 to J-1). The starting fields for the initialization of the forecast are therefore taken as the instantaneous field at 12:00 a.m. of Tuesday (J) resulting from the chain of daily analyses done each week for the previous 15 days.

The assimilation cycle is daily and all the data sets are assimilated intermittently at the end of each day after misfits (differences between forecast first guesses and observations) are computed at the time and location of the observations.

The MFS final products are daily mean fields of temperature, salinity, three-dimensional velocity field and sea level.

The preparation and run of the forecast is done through an automatic procedure which has been set up and tested during the MFSTEP project. The operational chain is activated as soon as the ECMWF forcing, the daily satellite SST and the SLA along-track data are available. The ten-day forecast fields and the last seven days of analysis are disseminated through an ftp site as soon as they are produced. A web bulletin is published every Tuesday on a dedicated web site (http://gnoo.bo.ingv.it/mfs). It is composed of four parts. (1) Maps of the position and of the vertical profiles of assimilated XBT and Argo data, maps of the along track assimilated SLA values and the animation of the last seven days of daily SST from satellite. (2) Maps of analysis fields. (3) Maps of forecast fields, such as the sea level, temperature, salinity, and horizontal velocities. (4) Analyses quality indices evaluated from the misfit values from the satellite SLA and the temperature and salinity profiles from XBT and Argo. The web publication of the bulletin is the last step in the operational chain. Usually, the whole procedure finishes on Wednesday morning, which means that there is a delay of less than 24 h in the forecast release.

The MFS data used in this study are the mean daily forecasts produced once a week from 16 August 2005 to 15 August 2006, a total of 53 forecasts, and the daily mean analyses. The study period has been chosen in order to consider data produced from the same version of the MFS system (version Sys2a).

3   Observational data and methods

The observations used in this work are the independent data from moored buoys and the quasi-independent observations from Argo and XBT which are assimilated by the system and described respectively in Poulain et al. (2007) and Manzella et al. (2007).

The moored buoys data are from the Puertos del Estado deep sea network (Alvarez-Fanjul et al., 2003), located where the water column is at least 200 m deep, and from the M3A network (Nittis et al., 2007; Nittis et al., 2003). The Puertos del Estado buoys are located close to: Alborán (36°16′01″ N 05°01′58″ E), Cabo de Gata (36°34′08″ N 02°19′12″ W), Cabo de Palos (37°39′03″ N 00°19′37″ W), Valencia (39°30′57″ N 00°12′14″ E or 39°27′43″ N 00°15′43″ E, depending on the period) and Tarragona (40°44′42″ N 01°27′25″ E), as shown in Fig. 2a, top panel. The E1-M3A buoy is located in the deep Cretan Sea (35°39′ N 24°58′ E, see Fig. 2a) and the W1-M3A in the Ligurian Sea (43°48′05″ N, 09°09′13″ E, see Fig. 2a), in an area approximately 1270 m deep. Only the surface measurements from all these buoys have been used to validate the MFS analyses since subsurface observations are too scarce in time.
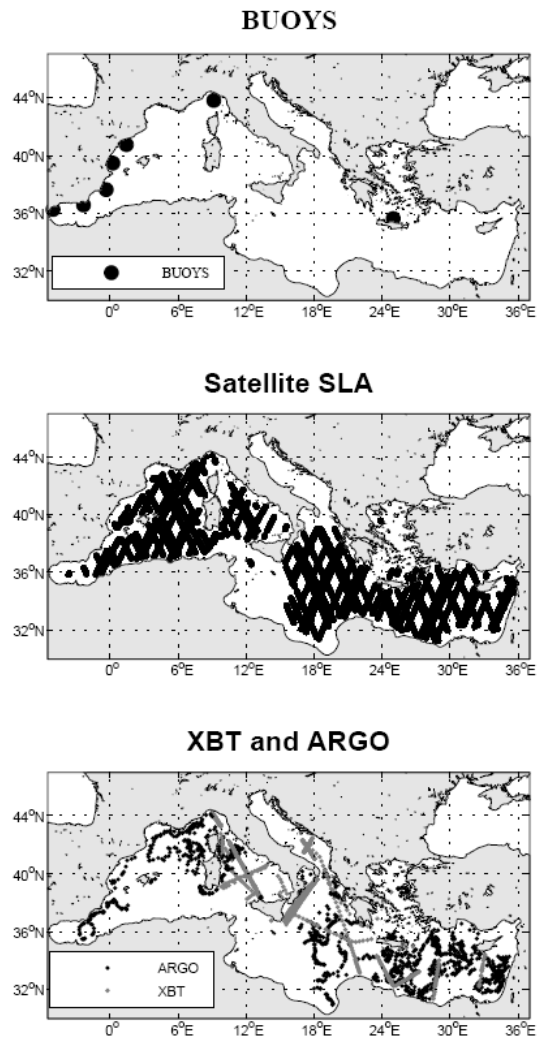

Fig. 2a. Maps of the observations available for the period of study. Upper panel: independent data from buoys; middle panel: quasi-independent data from SLA; bottom panel: quasi-independent XBT (grey) and Argo (black) profile locations.

As already said in the introduction, we evaluate the quality of the analyses and then we will use the latter to assess the forecast quality. The analyses data have been interpolated at each observational positions considering the four closest model grid points. The observations have been averaged daily before the inter-comparison and days with less then 3 observations are not considered.

In this paper we use the term rmse only with the differences between analyses and observations, computed as follows:

$$\text{rmse} = \sqrt{\frac{\sum_{1}^{N_{\text{obs}}} (X_{\text{obs}} - X_{\text{AN}})^2}{N_{\text{obs}}}} \qquad (1)$$

where X_{obs} is the observed value of Salinity or Temperature, X_{AN} the daily mean value of the Salinity or Temperature from analyses, interpolated into the observational position, and N_{obs} is the number of considered observations.

The forecast quality is assessed by computing differences with analyses. The differences between analyses and forecasts are due to:

– the assimilation of SLA, XBT and Argo data;

– the atmospheric forcing, since ocean forecasts are forced by atmospheric forecast while ocean analyses use atmospheric analysis fields;

– the assimilation of satellite SST during the analysis cycle.

The rmse of difference between the analysis and the forecast, so-called AF, is computed as follow:

$$\mathrm{AF}(t) = \left\langle \sqrt{\frac{\sum\limits_{1}^{N}(X_{\mathrm{AN}}(t) - X_{\mathrm{FC}}(t))}{N}} \right\rangle \tag{2}$$

where $X_{\mathrm{FC}}(t)$ is the daily mean values of temperature or salinity from the forecast and analysis respectively at each forecast day t, with t=d1, d2,..., d10, N are the 53 forecasts considered and the brackets indicate the normalised horizontal average at the selected depth (normalisation is the division by the area).

The rmse of the difference between analyses and persistence, so-called AP, and the difference between forecast and persistence, so-called FP, is computed as follow:

$$\mathrm{AP}(t) = \left\langle \sqrt{\frac{\sum\limits_{1}^{N}(X_{\mathrm{AN}}(t) - X_{\mathrm{AN}}(t=\mathrm{d}1))^2}{N}} \right\rangle \tag{3}$$

$$\mathrm{FP}(t) = \left\langle \sqrt{\frac{\sum\limits_{1}^{N}(X_{\mathrm{FC}}(t) - X_{\mathrm{AN}}(t=\mathrm{d}1))^2}{N}} \right\rangle \tag{4}$$

where the persistence (P) is considered to be the analysis mean daily field at the first day of forecast t=d1. In the atmospheric forecasting community, persistence is normally defined as the forecast initial condition, i.e., the analysis instantaneous field at time t=0. Here however we assess the quality of the daily mean forecast fields and we find it more appropriate to compare with the daily mean field analysis for the first day of the forecast.

In this study we have computed the AF, AP and FP for the Temperature and Salinity fields at selected levels of depth: 5 m, 30 m, 150 m, 300 m and 600 m. Following Murphy et al. (1988) the following skill score (SS) has been defined:

$$\mathrm{SS}(t) = \left(1 - \frac{\mathrm{AF}(t)}{\mathrm{AP}(t)}\right) * 100 \tag{5}$$

SS is equal to 100% if AF is equal to zero, which means a perfect forecast. Otherwise, SS is equal to 0 if AF is equal to AP, that means a very poor forecast. If AF value is less than AP, which means a forecast error less than persistence, then SS>0, i.e., positive SS values correspond to a gain of the forecast with respect to persistence. The contrary holds for SS<0.

The Anomaly Correlation, so-called AC, has been defined as:

$$\mathrm{AC}(t) = \left\langle \frac{\sum\limits_{1}^{N}\left(X^*_{\mathrm{AN}}(t) - \overline{X^*_{\mathrm{AN}}(t)}\right)\left(X^*_{\mathrm{FC}}(t) - \overline{X^*_{\mathrm{FC}}(t)}\right)}{\sqrt{\sum\limits_{1}^{N}\left(X^*_{\mathrm{AN}}(t) - \overline{X^*_{\mathrm{AN}}(t)}\right)^2 \sum\limits_{1}^{N}\left(X^*_{\mathrm{FC}}(t) - \overline{X^*_{\mathrm{FC}}(t)}\right)^2}} \right\rangle \tag{6}$$

where all the symbols have been explained before and $X^*_{\mathrm{AN}}(t)$ and $X^*_{\mathrm{FC}}(t)$ are defined as follow:

$$X^*_{\mathrm{AN}}(t) = X_{\mathrm{AN}}(t) - \overline{X_{\mathrm{AN}}(t)}$$

$$X^*_{\mathrm{FC}}(t) = X_{\mathrm{FC}}(t) - \overline{X_{\mathrm{FC}}(t)}$$

The over line indicate a temporal average which in our case we have chosen to be the daily mean seasonal value taken over July, August and September for Summer; October, November and December for Autumn; January, February and March for Winter and April, May and June for Spring. We know that the seasonal variations of the surface temperature in the ocean are more than 10°C and 0.1–0.2 psu with an almost periodic signal every year. This deterministic component is subtracted from the AC values in order to show the skill at the higher frequency variability, due to mesoscales and sub-basin scale gyres fluctuations (Pinardi et al., 2005).

Finally the bias of the difference between forecast and analysis has also been computed, i.e.:

$$\mathrm{Bias}(t) = \left\langle \frac{\sum\limits_{1}^{N}(X_{\mathrm{FC}}(t) - X_{\mathrm{AN}}(t))}{N} \right\rangle \tag{7}$$

where all the notation has been defined above.

4 Analysis and forecast evaluation

4.1 Rmse of observations and analyses

The rmse defined in Eq. (1) has been computed between the independent observations and the analyses for the temperature and salinity for all the days with at least three observations from buoys. The rmse for temperature could not therefore be computed from 8 May to 29 June 2006; 15, 16 and 27 July and from 31 July 31 to 15 August 2006. The rmse for salinity could not be computed from 16 August to 15 September 2005; 24–30 March 2006; from 7 May to 17 July 2006 and from 30 July 30 to 15 August 2006. The upper panels of Fig. 2b and c show the rmse for temperature and
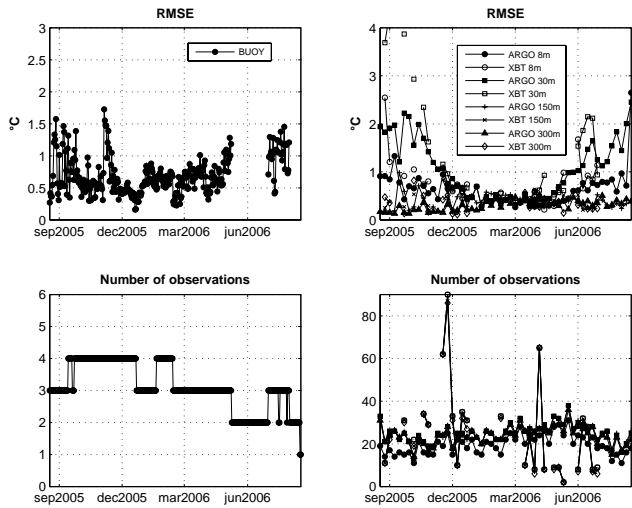
Fig. 2b. Independent and quasi-independent observations and comparison with analyses: rmse (Eq. 1) of temperature at surface (left-top panel) and the corresponding number of independent observations (left-bottom panel). Rmse of temperature for quasi-independent data (ARGO and XBT) at 8, 30, 150 and 300 m (right-top panel) and the corresponding number of observations (right-bottom panel). The round filled mark is the ARGO at 8 m, the round mark is the XBT at 8 m, the square filled mark is the ARGO at 30 m, the square mark is the XBT at 30 m, the + is the ARGO at 150 m, the x is the XBT at 150 m, the triangle is the ARGO at 300 m and the rhombus the XBT at 300 m.

salinity respectively. The number of measurements for each day is described in the corresponding bottom panels. For the quasi-independent data, data at the surface are available only starting from 8 m due to the pre-processing of this data (see Appendix A).

The number of observations in total however is very low; the independent data are around 4–3 buoys and the quasi-independent data around 20–30 profiles most of the days considered. Moreover, looking at the distribution of the buoys (Fig. 2a), is clear that most of the them are along the northern coast of the west basin. Notwithstanding all these limitations, the information provided from this evaluation gives an idea of the accuracy of the analysis of the MFS system.

The rmse mean value over the whole year is given in Table 1 for the various depths. The temperature rmse from the independent data and semi-independent data is comparable at the surface and at 8 m: 0.7°C for the independent and 0.8° for XBT and 0.6°C for Argo. The rmse at 30 m have the highest values from both Argo and XBT because this is the level at which the seasonal thermocline gradient peaks. The temperature rmse from quasi-independent data has (Fig. 2b, right panel) a clear seasonal cycle with higher values during the summer and lower values in winter, when the water column is almost completely unstratified. The seasonal signal is not clear for the independent observations, even though the higher values are in the summer period and this is prob-
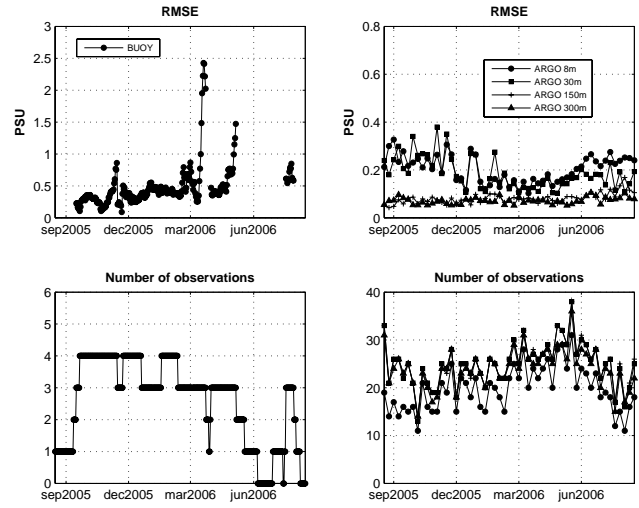
Fig. 2c. Independent and quasi-independent observations and comparison with analyses: rmse (Eq. 1) of salinity at surface (left-top panel) and the corresponding number of independent observations (left-bottom panel). Rmse of salinity for quasi-independent data (ARGO) at 8, 30, 150 and 300 m (right-top panel) and the corresponding number of observations (right-bottom panel). The round mark is for 8 m, the square for 30 m, the + for 150 m and the triangle for 300 m.

Table 1. Mean values of rmse AN-buoy, AN-Argo and AN-XBT at surface and different depths. The mean is computed over the considered year.

Depth	BUOY		ARGO		XBT
	T [°C]	S [psu]	T [°C]	S [psu]	T [°C]
Surface	0.7	0.4			
8 m			0.60	0.20	0.80
30 m			1.00	0.18	1.40
150 m			0.38	0.08	0.38
300 m			0.30	0.07	0.34

ably due to the fact that we are comparing values at surface and not in the thermocline. The salinity rmse at the surface is only estimated from independent observations and it has a mean value of 0.4 psu. There are two peaks of high rmse values at the end of March and in May 2006. The observations available for this period are from the buoys of Puertos del Estado located at Valencia, Capo de Gata and Terragona. In March–April 2006 the measures of the buoys of Valencia and Capo de Gata are fresher than the MFS model while the buoy of Terragona has a higher value of salinity respect the MFS model. The situation in May is exactly the opposite. The values of the difference between the buoy and the model is never higher than 1 psu except in March–April at the buoy of Valencia. Therefore these two peaks of high rmse could be due to the fact these buoys measure the salinity of the inflowing Atlantic water which forms large gradients with

the Mediterranean Sea waters. A small misplacement in the model anticyclonic gyres typical of this area could lead to very high rmse values. For the semi-independent data the salinity rmse mean value at 8 m is 0.2 psu and it decreases with depth. The seasonal signal in rmse is again evident for the quasi-independent observations at 8 m and 30 m (right panel of Fig. 2c).

The salinity and temperature at Capo de Gata has been compared with the model analysis, and the satellite SST as shown in Fig. 2d (upper panel). The satellite and the buoy measure the same temperature except few periods. In general the analyses are warmer then the buoy in spring and summer and colder in winter and autumn. This is in accordance with our previously discussed result of a positive bias due to summer excessive warming. The difference between the analysis and the buoy is never higher then 1°C during all the considered year, except for few days in summer time. The difference between analysis and satellite temperature shows that the relaxation through surface fluxes to satellite SST (Pinardi et al., 2003) in not affective parameterization of SST assimilation and it should modified in the future. The salinity of the buoy and of the analysis are very similar, with the exception of few days. The buoy in general is saltier then the analysis except in two episode in March and May when the salinity measured by the buoy decreases strongly and reaches values below 36 psu. This could be explain by the advective processes known to occur at position of the buoy, close to Almeria. There is a well known front (Millot, 1999) close to that area which crosses the Mediterranean from the Spanish to the Tunisian coast, the so called Almeria-Oran front characterized by a strong gradient of salinity between the incoming Atlantic water with a low salinity and the salty Mediterranean water. The location of this front is influenced by the anticyclonic gyres of the Alboran sea (Heburn et al., 1990), which varies in time and space. A misplacement in the model of the front, due to eddies phase errors, could be responsible of the major difference between the analysis and the buoy. From a recent work (Oddo et al., 2009) that studied the difference between our model and a new one with open lateral boundary conditions in the Atlantic, it has been documented that our model underestimates the salinity of the inflowing Atlantic water. This could explain why in general the analysis has a lower salinity than the buoy in this area.

4.2 Forecast evaluation

Before starting the discussion of the forecast skill scores, it is important to point out that the assumption that the analysis is the best estimate of the reality instead of an independent observation is mainly due to the fact that the aim of this work is to evaluate the forecast over all the Mediterranean basin and not only in those few positions where data are available. Even if it will not be possible to compute the absolute forecast error, we can try to estimate whether the ocean dynamics
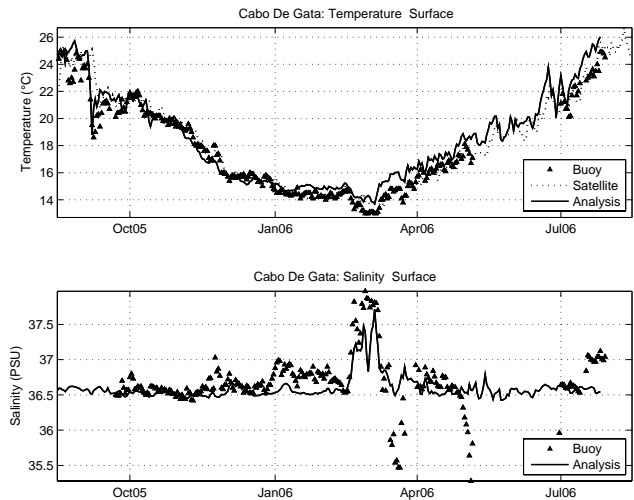

Fig. 2d. Temperature (upper panel) and salinity (bottom panel) time series at the buoys of Cabo Gata. The continuous line is the analysis, the triangle the buoy and for the temperature the dotted line is the OI-SST from satellite.

is affected over the forecast ten days by the atmospheric forcing inaccuracies and the assimilation.

In Fig. 3a we show the AF, AP and FP for temperature and salinity and for all the Mediterranean basin. The values of AP, FP are always bigger then AF except for AP on the first day (d1), due to the definition of P. This means that the system does better than persistence and it therefore makes sense to produce a ten-day forecast. In general, AP and FP provide information about the variability of the system in the ten days of the forecast for analysis and forecast. As expected, the differences between AP and FP are small and both increase rapidly, especially at the depths of 5 and 30 m, indicating that forecast and analysis have the same type of variability with respect to persistence. The only noticeable difference between AP and FP is at 5 m for temperature and this is due to differences in the atmospheric forcing used in the analysis and forecast. The latter in fact uses atmospheric forcing fields from forecasts while the former uses the atmospheric analysed surface fields.

At the last day of forecast (d10), FP and AP for temperature at 5 m have a value that is double of AF and 30% higher than AF for salinity. At 30 m, both salinity and temperature FP and AP are 30% higher than AF after the first four days. The values of AF and FP down to 150 m are very small, especially for temperature, due to the fact that at depth the ocean variability is much less than at the surface. The values of AF at the different depth have been compared with the variability of the system as expected by the annual average of error standard deviation estimated from the background error covariance matrix of the OI scheme. Figure 3b shows the averaged background error covariance (as described by Dobricic et al., 2005) at the selected depth of this study (5, 30, 150, 300 and
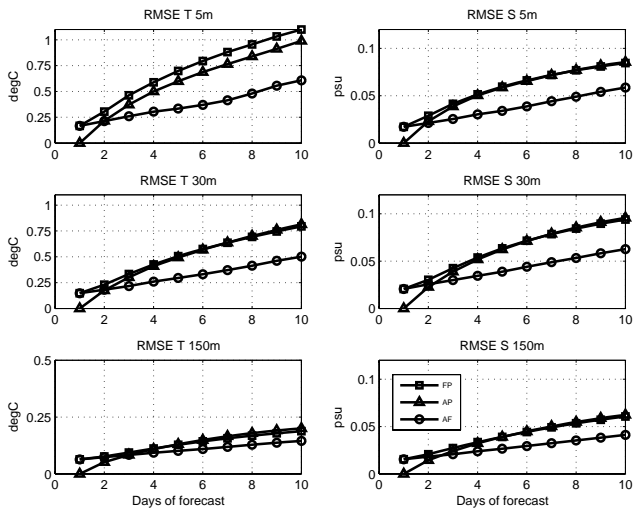
Fig. 3a. Rms of Analysis-Persistence (AP, Eq. 3), Forecast-Persistence (FP, Eq. 4) and Analysis-Forecast-(AF, Eq. 2) for Temperature (left panel) and Salinity (right panel) at the depths of 5, 30 and 150 m. The rmse is a mean of the 53 ten-day forecast cycles from August 2005 to 2006.

Fig. 3b. Vertical profile of spatially and annual averaged of error standard deviation estimated from the background error covariances of the OI scheme and rmse of Analysis-Forecast (AF) for Temperature (top panel) and Salinity (bottom panel).

600 m) for both temperature (upper panel) and salinity (lower panel) compared with the rmse of analysis-forecast, AF, already shown in Fig. 3a. The values of the Temperature rmse at 5 and 30 m are lower than the OI error variance estimated from the system variability until the seventh-eighth forecast day. The salinity rmse is always lower then the expected variance of the field because this field has a slower variability in time respect the ten days of forecast considered here.

The differences between analyses and forecasts are partly due to the different atmospheric forcing and to the assimilation. It is not easy to quantify the impact of each of these components using the available operational products, without performing ad hoc experiments. It is however possible to estimate differences between atmospheric forcing used to produce the analysis and the forecast fields (ECMWF analysis and forecast respectively) and the impact they have on the model heat and momentum fluxes. The upper panel of Fig. 4 shows the normalized rmse difference between the analyses and the forecast fields for all the parameters used to force the ocean model (Mean Sea Level pressure, m.s.l.; air temperature at 2 m, T2M; dew point temperature at 2 m, TD2M; zonal wind speed at 10 m, U10M; meridional wind speed at 10 m, V10M; cloud cover, CLC) and for the ten days of the forecast. It is clear from Fig. 4 that the uncertainty grows with time in the atmospheric forcing in a rather similar way of the growth of errors in the oceanic fields (Fig. 3). The AF for model heat and momentum fluxes are shown in the bottom panels of Fig. 4 and again they show an almost linear increase from the first forecast day to the tenth. Thus it is reasonable to think that part of the AF growth in Fig. 3 is due to the growth of uncertainties in the atmospheric forcing for the ocean but we cannot quantify how much.
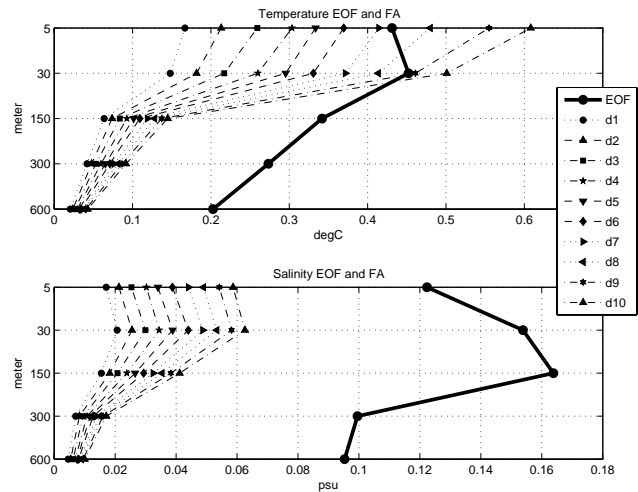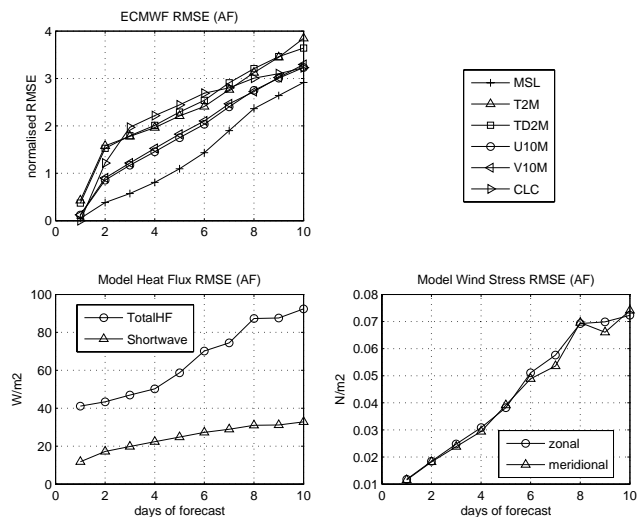

Fig. 4. Upper Panel: normalised rms of the difference between ECMWF atmospheric forecast and analyses fields. The value is a mean of the 53 ten-day forecast cycles. The value of the rms is normalised by the standard deviation of each atmospheric field considered. The atmospheric fields considered are: Mean Sea Level Pressure (m.s.l.), Air Temperature at 2 m (T2M), Dew-point Temperature at 2 m (T2MD), zonal and meridional wind speed at 10 m (U10M and V10M) and the Cloud Cover (CLC). Bottom Panel: AF of the model heat fluxes (total upward heat flux and shortwave radiation flux) and momentum (meridional and zonal wind stress).

In order to try to estimate the percentage of improvement in the accuracy of the forecast, SS has been computed from AF and AP. Figure 5 shows the ten days of SS for the 5 m, 30 m and 150 m. The first day of forecast SS could not be computed because AP is equal to zero. SS is always positive
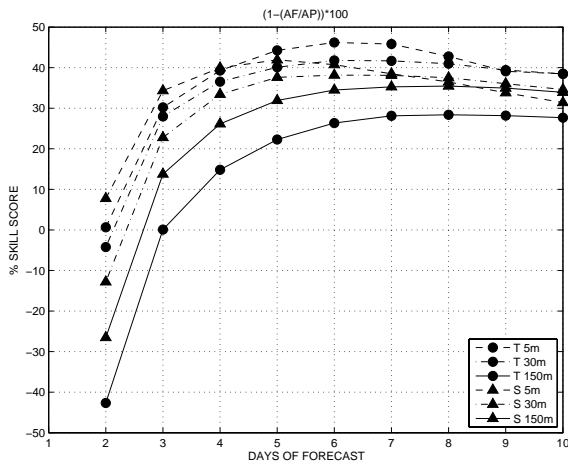
Fig. 5. Skill Score (SS, Eq. 5) for salinity (triangle) and Temperature (circle) at the depths of 5 m (dashed line), 30 m (dash dotted line) and 150 m (continuos line). The SS is a mean of the 53 ten-day forecast cycles from August 2005 to 2006.

Fig. 6. Bias for the difference between forecast and analysis (Eq. 6) of Temperature (top panel) and Salinity (bottom panel) at 5, 30, 150, 300 and 600 m. The profiles for the ten days of forecast are labelled as d1, d2, d3, d4, d5, d6, d7, d8, d9 and d10 in the corresponding legend.

for salinity and temperature at 5 m, reaching values of 45% at the 5th-6th day of forecast, which is the maximum value. After this point, the values start to decrease a little, but in any case maintain values around 30–40%. This means that the improvement in the accuracy of the forecast with respect to persistence is high for the first days of the forecast. SS is negative for the second day at 30 and 150 m. While at 5 m the behaviour of SS is approximately the same for temperature and salinity, at 150 m depth the salinity SS has much higher values that that of temperature. The salinity SS shows a behaviour close to that for 5 and 30 m, while the temperature SS always has lower values. The temperature SS at 150 m reaches a positive value only at the third day and the maximum is 20% at the 7th day of forecast.

Thus the first conclusion is that while at the surface the forecast is better than persistence especially in the first three days, down to 150 m the persistence does better in the first days. This is related to the ocean dynamics which has faster processes at the surface, in the mixed layer and slower, geostrophic motion below it. Forecast starting from the assimilated initial condition may suffer of adjustment due to the data insertion and thus produce worse conditions at the beginning. The time scale of several days seems to be necessary in the subsurface to obtain SS values of the order of 30%.

The bias of the difference between forecast and analyses (Eq. 7) is shown in Fig. 6a for temperature (top panel) and salinity (bottom panel). The bias is increasing from d1 to d10, as might be expected. For the temperature, the bias is positive at 5 m (with values from +0.02°C at d1 to ca. +0.1°C at d10) and negative at all the other depths. The values at 30 m are of −0.02°C at d1, and of −0.06°C at d10. The bias is very small from 150 m down to 600 m. The forecast is therefore warmer at surface and colder at depth. We ar-
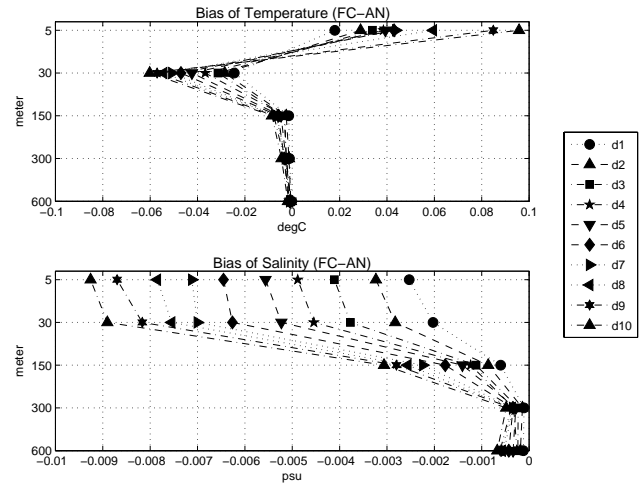
gue that the reason for the opposite sign bias between 5 and 30 m is due to the higher water column stratification due to the surface positive bias that forces the heat not to be transferred downward, allowing for the negative bias to increase at 30 m. This error is than a seasonally dependent error due to a combination of surface fluxes and radiative heat penetration errors during the summer. The model therefore does not reproduce correctly the depth of the seasonal thermocline (Tonani et al., 2008). The bias of salinity, however, is always negative, with decreasing values from surface to depth. The values are from −0.02–0.025 psu at d1 at 5 and 30 m and around −0.01 psu at d10, therefore the forecast is always fresher than the analysis. The model water fluxes in our model (Tonani et al., 2008) induce lower entering Atlantic water salinities, as documented by Oddo et al. (2009) and this generally produces a lower salinity signal in the while water column. An additional reason for the lower salinities is also that the evaporation of our model is not large enough in the Eastern Mediterranean and the Levantine Intermediate Water formed there, after summer evaporation, are not salty enough.

The Anomaly Correlation (AC) is shown in Fig. 7 for temperature (left panel) and salinity (right panel). The AC has values that decrease from the first day to the last reaching the minimum value of 0.78 at d10. The AC of temperature at surface (depth of 5 and 30 m) has higher values than AC at depth. This could be hinting to the fact that atmospheric forcing inaccuracies have smaller impact on the forecast accuracy than non dynamically adjusted corrections due to data assimilation in the initial condition.
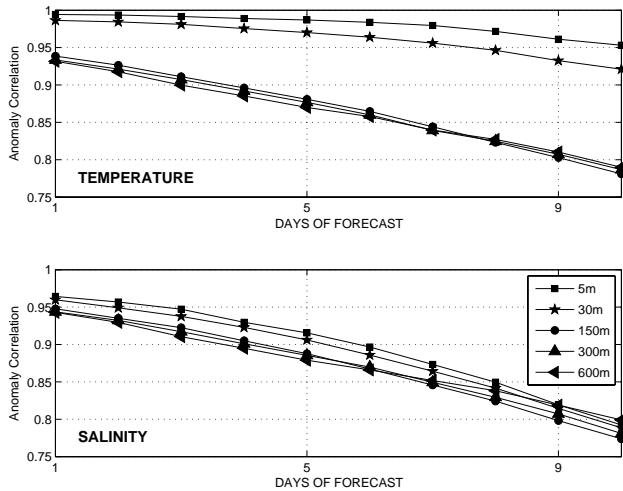
Fig. 7. Anomaly correlation (AC, Eq. 5) of Temperature (left panel) and Salinity (right panel) at the depths of 5 m, 30, 150 m, 300 m and 600 m.

Fig. 8a. Map of the absolute value of the difference between forecast and analysis at the fifth day of forecast at 30 m for salinity (top panel) and temperature (bottom panel). The absolute value of the forecast-analysis is the mean of the 53 ten-day forecast cycles from August 2005 to 2006.

We now consider how the quality of the forecast varies in the different regions of the Mediterranean Sea. This is not simple because we should take into account the fact that the quality of the analysis can vary from area to area depending on the spatial distribution of the observations and the different dynamics. Figure 8a shows the map of the absolute value of the difference between analyses and forecast for the fifth day for forecast (d5) at 30 m for both salinity and temperature. The comparison with Fig. 2a shows clearly that where the system does not assimilate observations the difference between analysis and forecast is low. This is a limit of our assumption that considers the analysis the truth In order to estimate the sensitivity of our method to the availability of observations in the analysis, we followed the advice to represent the difference between forecast analysis only at points where there are observations. We have recomputed the values represented in Fig. 3 only using points where the absolute value of the difference between the forecast and analysis is greater then 0.15°C for the temperature and 0.025 psu for the salinity. This result are shown in Fig. 1 review below where the values of FA are compared with and without the masking of the points with differences less than the threshold. The areas masked out with this criteria coincide with the region void of observations. The rmse threshold has been established over the statistics on the absolute value of the differences between analysis and forecast at the fifth day of forecast (d5) for the studied period (Fig. 8a). Figure 8b shows the rmse computed on the reduced number of grid point for AP (rmse of analysis-persistence), FP (rmse of forecast-persistence) and AF (rmse of analysis-forecast). The rmse is slightly higher then the values shown in Fig. 3 but the error growth is very similar and in all cases the forecast wins over persistence.
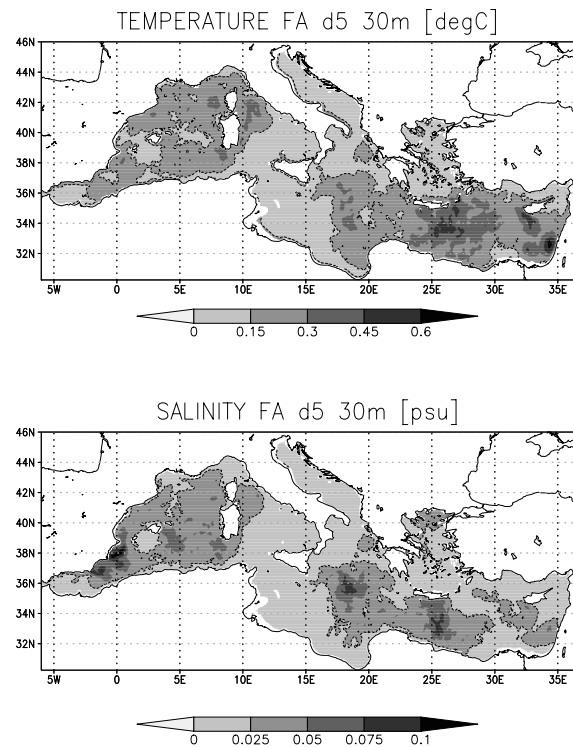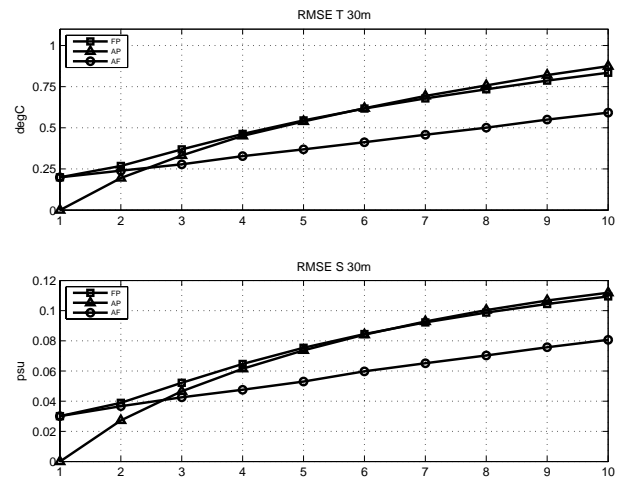

Fig. 8b. Rms of Analysis-Persistence (AP, Eq. 3), Forecast-Persistence (FP, Eq. 4) and Analysis-Forecast-(AF, Eq. 2) for Temperature (top panel) and Salinity (bottom panel) at the depths of 30 m. The rmse is a mean of the 53 ten-day forecast cycles from August 2005 to 2006 computed only in the region identified by Fig. 8a. The temperature rmse have been computed only where the value of FA is higher then 0.15°C, the salinity rmse is computed only in the areas where FA is higher then 0.025 psu.
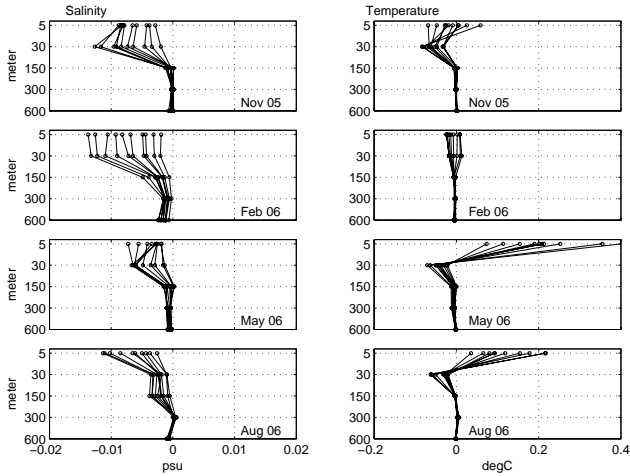
Fig. 9a. Monthly mean bias for selected month for Salinity (left panel) and Temperature (right panel). The bias is computed at the depths of 5, 30 , 150, 300 and 600 m for the months of November 2005, February 2006, May 2006 and August 2006.

The monthly variability of bias and AP has been computed in order to try to understand better how the method we have described for the forecast evaluation is related to the seasonal variability of the system. Figure 9a shows the vertical profiles of the monthly mean bias of Forecast-Analysis for the selected months of November 2005, February, May and August 2006. The left panel represents the bias of salinity and the right panel for temperature. The seasonal variability of the bias is evident especially for temperature. The bias is between 0 and 0.1°C and it is generally negative in the subsurface while it is positive at the surface during the summer months. The reason for that was studied in the past by Byun and Pinardi (2007). This is due to the overestimation of the solar incoming radiation with in our model (an approximate form of the Smithsonian formula modified by Reed, 1977). Figure 9b represents the monthly mean surface temperature from the model analyses and from the Satellite SST. The analyses are colder than satellite in autumn and warmer in spring-summer. As we have explained in one of the previous points, there is a seasonally dependent error due to a combination of surface fluxes and radiative heat penetration during the summer.

Figure 10 shows the AP at 30 m for the temperature fields for the selected months of 5 November, 6 February, 6 May and 6 August. It is clear that the largest variability is in November and August with respect to February and May. The same behaviour, even though less evident, is also shown by salinity (not shown). This is also reflected in the SS computed for the corresponding months (Fig. 11). SS has a higher score during spring, summer and autumn and a lower during winter, which means that this score is higher when the variability of the system is bigger. This means that when the
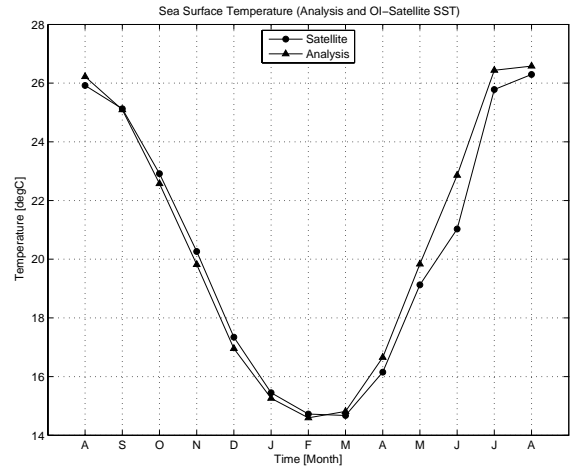

Fig. 9b. Monthly mean sea of Sea Surface Temperature from Analysis and satellite OI-SST.
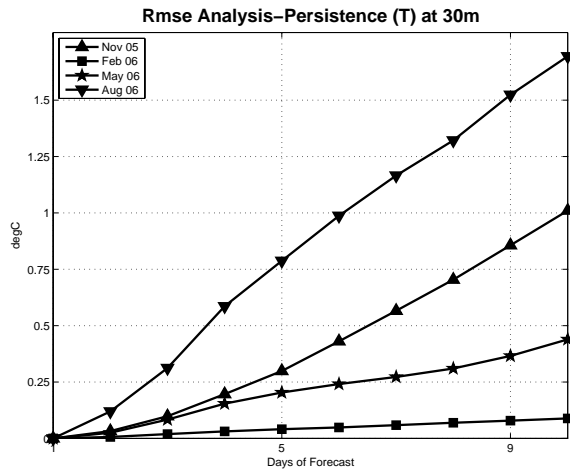

Fig. 10. Monthly mean of AP for Temperature at 30 m for November 2005, February 2006, May 2006 and August 2006.

system has a high variability the forecast has more skill with respect to the persistence.

5   Conclusions

This work is the first attempt to estimate the quality of the forecast produced during the MFSTEP project and will be useful in the future to define the forecast evaluation protocol. There is early evidence that it is important to find out a way to weight the quality of the analysis with respect to the number of data assimilated and to understand better the impact of the data assimilation corrections at depth.

The rmse confirms that the quality of the analyses is reasonable and that errors converge in the interval 0.4–1°C for temperature and 0.4–0.07 psu for salinity. Largest errors appear at the thermocline depth and they are largely seasonal.
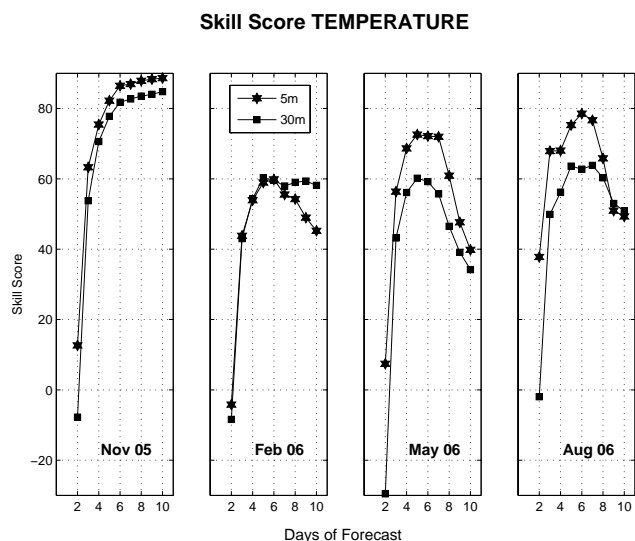
Skill Score TEMPERATURE

Fig. 11. Monthly mean of SS for Temperature at 5 and 30 m for November 2005, February 2006, May 2006 and August 2006.

The MFSTEP forecast for the period 2005–2006 is always, except at 150 m, better than persistence. This result is important to justify the effort that has to be made to set up and maintain a forecasting system.

High values of SS up to 45% occur for the surface layer for both salinity and temperature while SS is less than zero at depth for several days. The growth of errors in the oceanic state variables is due both to atmospheric forcing uncertainties and to model errors related to missing and/or not well modelled processes.This is a basin-scale assessment which will be followed by work on sub-regional assessment.

Appendix A

Assimilated data

The data collected every week on Tuesday by the forecasting system are: satellite data of SLA and SST, in-situ profiles from Argo and XBT and the atmospheric forcing. The analyses and forecast of the meteorological fields are from the European Centre for Medium-Range Weather Forecasts (ECMWF) and have a time resolution of 6hr and the horizontal grid has a resolution of $0.5° \times 0.5°$ ca. The fields are: Cloud Cover, Air Temperature at 2 m, zonal and meridional wind speed at 20 m, dew point temperature.

The MFS SST product is based on night-time AVHRR images acquired and processed by the Gruppo di Oceanografia da Satellite, Institute of Atmospheric Sciences and Climate of the Italian National Research Council (GOS-CNR-ISAC). In particular, mapped AVHRR SST fields are computed only with night-time passages of NOAA-AVHRR-14 and NOAA-

AVHRR-15 satellite (Marullo et al., 2007; Buongiorno et al., 2003).

The SLA data are collected by Collection et Localisation Satellitaire (CLS) in Toulouse. MFSTEP uses the along-track product of two missions: Jason1 and Geosat Follow On (GFO). Jason1 has a repeat time of ten days, and the data are routinely produced by CNES/NASA. Geoast Follow On (GFO) data, which have a repeat time of 17 days, are provided by NOAA. The products delivered are computed with precise Orbit Error Reduction and Long Wavelength Error reduction (OER products).

The vertical temperature profiles are collected along several routes in the Mediterranean Sea. The near real time quality control is done by ENEA and the procedure is described in Manzella et al. (2003). These data are downloaded from ENEA server in La Spezia. The vertical profiles of XBT are interpolated at the MFS model vertical levels before being assimilated into the system.

The Argo vertical temperature and salinity profiles are collected by ARGO floats from MedArgo (Poulain et al., 2007). The data collection centre and first quality control is done at the Coriolis data centre in IFREMER (Brest). The INGV pre-processing procedure consist of two steps. First the profile flags are checked and only high quality profiles are retained. In particular, the procedure checks: the flag for the date and position flag; the flag of each temperature and salinity profile; the pre-processing of the selected profiles. The second part of the pre-processing consists of the computation of the Brunt-Vaisala frequency. Where the Brunt-Vaisala is negative, the temperature and salinity data are rejected. The entire profile is rejected if the distance between the data points in the first 150 m is greater then 40 m.

References

Alvarez Fanjul, E., Alfonso, M., Ruiz, M. I., Lopez, J. D., and Rodriguez, I.: Real time monitoring of Spanish coastal waters: The deep water network", Proceedings of the Third International Conference on EuroGOOS, Athens, 2003.

Buongiorno Nardelli, B., Larnicol, G., DAcunzo, E., Santoleri, R., Marullo, S., and Le Traon, P. Y.: Near Real Time SLA and SST products during 2-years of MFS pilot project: processing, analysis of the variability and of the coupled patterns, Ann. Geophys., 21, 103–121, 2003,
http://www.ann-geophys.net/21/103/2003/.

Byun, D.-S. and Pinardi, N.: Comparison of Marine Insolation Estimating methods in the Adriatic Sea, Ocean Sci. J., 42(4), 211–222, 2007.

Demirov, E., Pinardi, N., Fratianni, C., Tonani, M., Giacomelli, L., and De Mey, P.: Assimilation scheme of the Mediterranean Forecasting System: operational implementation, Ann. Geophys., 21,

189–204, 2003,
http://www.ann-geophys.net/21/189/2003/.

Dobricic, S., Pinardi, N., Adani, M., Tonani, M., Fratianni, C., Bonazzi, A., and Fernandez, V.: Daily oceanographic analyses by Mediterranean Forecasting System at the basin scale, Ocean Sci., 3, 149–157, 2007,
http://www.ocean-sci.net/3/149/2007/.

Dobricic, S., Pinardi, N., Adani, M., Bonazzi, A., Fratianni, C., and Tonani, M.: Mediterranean Forecasting System: a new assimilation scheme for sea level anomaly and it's validation, Q. J. Roy. Meteorol. Soc., 128, 1–12, 2004.

Heburn G. W. and La Violette, P. E.: Variations in the structure of the anticyclonic gyres found in the Alboran Sea, J. Geophys. Res., 95(C2), 1599–1613, 1990.

Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification: A Practitioner's Guide in Atmospheric Science, Wiley & Sons Ltd, 2003.

Lermusiaux P. F. J.: Uncertainty estimation and prediction for interdisciplinary ocean dynamics, J. Comp. Phys., 217, 176–199, 2006.

Manzella, G. M. R., Reseghetti, F., Coppini, G., Borghini, M., Cruzado, A., Galli, C., Gertman, I., Gervais, T., Hayes, D., Millot, C., Murashkovsky, A., Özsoy, E., Tziavos, C., Velasquez, Z., and Zodiatis, G.: The improvements of the ships of opportunity program in MFS-TEP, Ocean Sci., 3, 245–258, 2007,
http://www.ocean-sci.net/3/245/2007/.

Manzella, G. M. R., Scoccimarro, E., Pinardi, N., and Tonani, M.: Improved near real-time data management procedures for the Mediterranean ocean Forecasting System-Voluntary Observing Ship program, Ann. Geophys., 21, 49–62, 2003,
http://www.ann-geophys.net/21/49/2003/.

Marullo, S., Buongiorno Nardelli, B., Guarracino, M., and Santoleri, R.: Observing the Mediterranean Sea from space: 21 years of Pathfinder-AVHRR sea surface temperatures (1985 to 2005): re-analysis and validation, Ocean Sci., 3, 299–310, 2007,
http://www.ocean-sci.net/3/299/2007/.

Millot, C.: Circulation in the Western Mediterranean Sea, J. Marine Syst., 20, 423–442, 1999.

Murphy, A. H.: Skill score based on the mean square error and their relation to the correlation coefficient, Mon. Weather Rev., 116, 2417–2424, 1988.

Murphy, A. H.: What is a good forecast? An essay on the nature of goodness in weather forecasting, Weather Forecast., 8, 281–293, 1993.

Nittis, K., Tziavos, C., Thanos, I., Drakopoulos, P., Cardin, V., Gacic, M., Petihakis, G., and Basana, R.: The Mediterranean Moored Multi-sensor Array (M3A): system development and initial results, Ann. Geophys., 21, 75–87, 2003,
http://www.ann-geophys.net/21/75/2003/.

Nittis, K., Tziavos, C., Bozzano, R., Cardin, V., Thanos, Y., Petihakis, G., Schiano, M. E., and Zanon, F.: The M3A multi-sensor buoy network of the Mediterranean Sea, Ocean Sci., 3, 229–243, 2007,
http://www.ocean-sci.net/3/229/2007/.

Oddo, P., Adani, M., Pinardi, N., Fratianni, C., Tonani, M., and Pettenuzzo, D.: A nested Atlantic-Mediterranean Sea general circulation model for operational forecasting, Ocean Sci., 5, 461–473, 2009,
http://www.ocean-sci.net/5/461/2009/.

Pinardi, N., Allen, I., Demirov, E., De Mey, P., Korres, G., Lascaratos, A., Le Traon, P.-Y., Maillard, C., Manzella, G., and Tziavos, C.: The Mediterranean ocean forecasting system: first phase of implementation (1998–2001), Ann. Geophys., 21, 3–20, 2003,
http://www.ann-geophys.net/21/3/2003/.

Pinardi, N., Arneri, E., Crise, A., Ravaioli, M., and Zavatarelli, M.: The physical, sedimentary and ecological structure and variability of shelf areas in the Mediterranean Sea, The Sea, Vol. 14, Harvard University Press, 1243–1330, 2005.

Poulain, P.-M., Barbanti, R., Font, J., Cruzado, A., Millot, C., Gertman, I., Griffa, A., Molcard, A., Rupolo, V., Le Bras, S., and Petit de la Villeon, L.: MedArgo: a drifting profiler program in the Mediterranean Sea, Ocean Sci., 3, 379–395, 2007,
http://www.ocean-sci.net/3/379/2007/.

Reed, R. K.: On estimating insolation over the ocean, J. Phys. Oceanogr., 7, 482–485, 1977.

Tonani, M., Pinardi, N., Dobricic, S., Pujol, I., and Fratianni, C.: A high-resolution free-surface model of the Mediterranean Sea, Ocean Sci., 4, 1–14, 2008,
http://www.ocean-sci.net/4/1/2008/.