



**HAL**  
open science

## Assimilation scheme of the Mediterranean Forecasting System: operational implementation

E. Demirov, N. Pinardi, C. Fratianni, M. Tonani, L. Giacomelli, P. de Mey

► **To cite this version:**

E. Demirov, N. Pinardi, C. Fratianni, M. Tonani, L. Giacomelli, et al.. Assimilation scheme of the Mediterranean Forecasting System: operational implementation. *Annales Geophysicae*, 2003, 21 (1), pp.189-204. hal-00329225

**HAL Id: hal-00329225**

**<https://hal.science/hal-00329225>**

Submitted on 18 Jun 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Assimilation scheme of the Mediterranean Forecasting System: operational implementation

E. Demirov<sup>1</sup>, N. Pinardi<sup>2</sup>, C. Fratianni<sup>1</sup>, M. Tonani<sup>1</sup>, L. Giacomelli<sup>2</sup>, and P. De Mey<sup>3</sup>

<sup>1</sup>Instituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

<sup>2</sup>Bologna University, Corso di Scienze Ambientali, Ravenna, Italy

<sup>3</sup>LEGOS, Toulouse, France

Received: 12 December 2001 – Revised: 11 June 2002 – Accepted: 8 July 2002

**Abstract.** This paper describes the operational implementation of the data assimilation scheme for the Mediterranean Forecasting System Pilot Project (MFSPP). The assimilation scheme, System for Ocean Forecast and Analysis (SOFA), is a reduced order Optimal Interpolation (OI) scheme. The order reduction is achieved by projection of the state vector into vertical Empirical Orthogonal Functions (EOF). The data assimilated are Sea Level Anomaly (SLA) and temperature profiles from Expandable Bathy Termographs (XBT). The data collection, quality control, assimilation and forecast procedures are all done in Near Real Time (NRT). The OI is used intermittently with an assimilation cycle of one week so that an analysis is produced once a week. The forecast is then done for ten days following the analysis day.

The root mean square (RMS) between the model forecast and the analysis (the forecast RMS) is below 0.7°C in the surface layers and below 0.2°C in the layers deeper than 200 m for all the ten forecast days. The RMS between forecast and initial condition (persistence RMS) is higher than forecast RMS after the first day. This means that the model improves forecast with respect to persistence. The calculation of the misfit between the forecast and the satellite data suggests that the model solution represents well the main space and time variability of the SLA except for a relatively short period of three – four weeks during the summer when the data show a fast transition between the cyclonic winter and anti-cyclonic summer regimes. This occurs in the surface layers that are not corrected by our assimilation scheme hypothesis. On the basis of the forecast skill scores analysis, conclusions are drawn about future improvements.

**Key words.** Oceanography; general (marginal and semi-enclosed seas; numerical modeling; ocean prediction)

## 1 Introduction

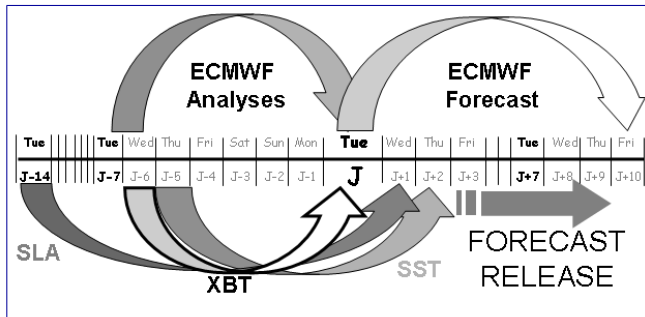
One major goal of the Mediterranean Forecasting System Pilot Project (MFSPP, Pinardi et al., 2003) was to demon-

strate the feasibility of operational predictions of the baroclinic basin scale circulation. Following the scientific plan of the project (see Pinardi and Flemming, 1998), a 10 days forecasting system of currents, temperature and salinity fields was set up starting from 4 January 2000. The forecasting system includes three major elements: a data collection network, the general circulation model and the data assimilation scheme. Presently, this forecasting system produces a weekly ten days basin scale forecast, which is published on a perpetual Web site (<http://www.cineca.it/mfspp>). In the future, the operational forecasting activities will also include downscaling towards the shelf areas with nested models and ecological forecasting (Pinardi et al., 2002); the basin scale forecasts will provide the initial and boundary conditions for higher resolution nested shelf model forecasts.

The main elements of the basin scale forecasting system are reviewed in detail in different papers. The observing system is described in Manzella et al. (2001) for the Voluntary Observing Ship (VOS), and Le Traon and Ogor (1998) for the satellite data. The ocean general circulation model (OGCM) of the basin scale forecasting system is presented in Demirov and Pinardi (2002). The data assimilation scheme is described and tested with a Mediterranean Sea circulation model by De Mey and Benkiran (2002) for satellite altimeter data and twin experiments XBTs.

In the present paper we describe the operational implementation of the data assimilation scheme in a multivariate mode, never tried before, and we analyze the skill of the forecast with respect to observations and analyses for a six month period during the Targeted Operational Period of MFSPP. This paper includes two main results: (a) the description of the method of assimilation of both SLA and XBT with one week assimilation cycle and (b) the evaluation of the forecast performance.

The use of the Mediterranean Sea OGCM, forced with surface meteorological fields in order to calculate heat and momentum fluxes at the air-sea interface, is discussed in Castellari et al. (1998), Castellari et al. (2000) and Demirov and Pinardi (2002). These previous simulations showed that the model is capable of reproducing the major features of the



**Fig. 1.** Schematic NRT data collection system in MFSP containing both oceanic and atmospheric data. The starting day of the forecast is indicated by  $J$ . Each week the system is repeated. Forecast is released Friday with three days delay.

circulation and water mass variability known from observations. However, there are still uncertainties in the model formulation that could produce incorrect forecasts. Between them, the relatively coarse resolution of the model in some areas (for instance in the Straits), uncertain subgrid scale physics parameterizations and rather modest spatial and temporal resolution of the surface forcing fields (half a degree and every six hours), could produce inaccuracies in the initial conditions for the forecasts.

The predictability limit is strongly dependent upon the quality of the nowcast that in turn depends on the observing system and the data assimilation scheme. Thus the development of an optimal system that assimilates the NRT observations was a major step during MFSP. The assimilation scheme first of all should have been multivariate both in input and output: the entering data are SLA and XBT profiles, but the corrections to the model first guess will be done on temperature, salinity and stream function, i.e. the baroclinic and barotropic components of the dynamical fields.

Last but not least, the forecast should be evaluated with respect to consistency, quality and accuracy (Murphy, 1993). This means that the predicted fields are objectively and subjectively compared to observations or analyses, trying to find out if and how much the model is capable of reproducing reality, how important is the data distribution and quality for the forecast, etc. The quality and accuracy is mainly given in terms of skill scores of various nature, already used in the meteorological literature (Murphy, 1988), and previous ocean forecasts (Walstad and Robinson, 1990).

The paper is organized in the following way. Section 2 discusses the data acquisition and organization of the forecast cycle. Section 3 presents the assimilation scheme. Section 4 discusses the results in terms of skill scores. Finally, Sect. 5 presents the conclusions.

## 2 The data acquisition and forecast cycle

The nowcast – forecast procedure of MFSP is shown in Fig. 1. Since 4 January 2000 the MFSP forecast is pro-

duced weekly with starting time  $J$ , which is Tuesday at noon of every week. The preparation and run of the forecast is done on several stages. Firstly, all data needed for the assimilation procedure and surface forcing calculation are collected. The procedure of data collection and quality check is briefly described in Appendix A. When the ocean satellite and insitu data and ECMWF analyses for the past 14 days are collected, the MFSP nowcast/forecast procedure is run for the past 14 days to produce the model initial condition for the starting day  $J$  of the forecast. The organization of the analysis run is related to the specific features of the MFSP data assimilation scheme and is described in detail in the next paragraph. When the model initial condition for day  $J$  is calculated, the forecast run is done for 10 days forward. The ocean general circulation model (OGCM), used in the analysis and forecast runs is described in Appendix B.

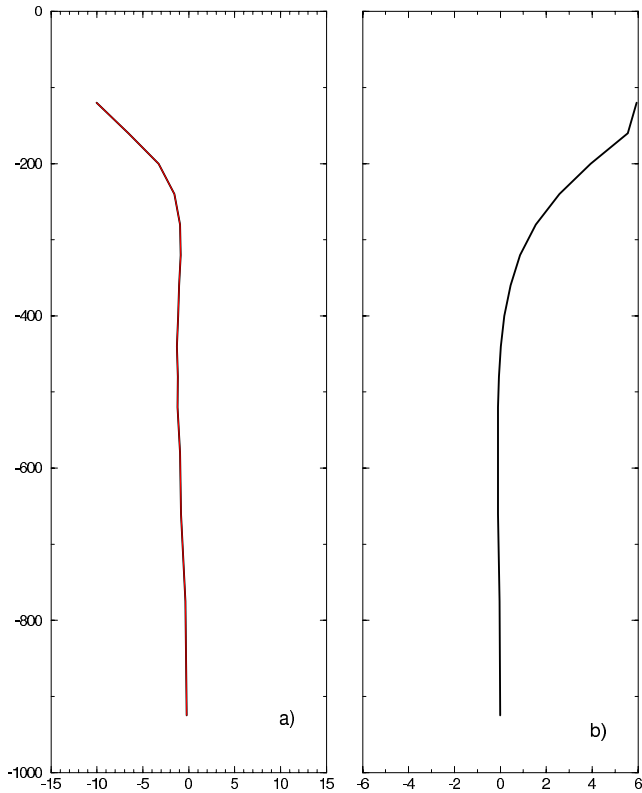
The ocean and atmospheric data used in the analysis and forecast are collected on the day  $J + 1$  (the Wednesday of every week), except the SST field, which only becomes available on the day  $J + 2$  (or Thursday of every week). Then the computational procedure requires between 6–10 h CPU time for the analysis and 1.5 h for the forecast. The CPU time for the analysis depends on the amount of assimilated ocean data, which can vary from one week to another. The forecast and analysis fields are published on the Web every Friday afternoon, i.e. with about 3 days delay after the starting day ( $J$ ) of the forecast. This delay is mainly due to the time needed for processing the ocean SST data. Presently, the satellite SST data become available on  $J + 1$  (Wednesday), which makes it possible to run the forecast with only two days delay, i.e. now the forecast becomes available on Thursday of every week.

## 3 The Data assimilation scheme

The data assimilation scheme used in MFSP is the System for Ocean Forecasting and Analysis (SOFA). The main definitions, notations and a brief description of the scheme are presented in Appendix C. A more detailed discussion of SOFA can be found in De Mey and Benkiran (2002).

SOFA is a reduced order multivariate optimal interpolation scheme. The order reduction is achieved by projecting the state vector onto vertical EOFs which are the eigenvectors of the error covariance matrix for the forecast. The scheme is multivariate in terms of data input and corrections made on the model solution. Our particular state vector contains the model predictive variables such as temperature, salinity and barotropic stream function. The existence of dominant EOFs' for the Mediterranean temperature-salinity relationship is discussed and demonstrated in Sparnocchia et al. (2003). Here, as in many other operational schemes, we use EOFs deduced from data rather than from an analysis of the forecast error covariance matrix.

De Mey and Robinson (1987) applied, for the first time, the order reduction in ocean data assimilation with a quasi-geostrophic model. They used a vertical EOF to project sur-



**Fig. 2.** Bivariate Empirical Orthogonal Functions used in assimilation of SLA: (a) Temperature EOF, (b) Salinity EOF.

face altimeter information, i.e. the surface quasigeostrophic stream function at different depths, and then initialize the forecasts. On the basis of this experience, the order reduction operator of SOFA is used here with vertical EOFs that this time are multivariate.

The observations do not need to be model dynamical variables but they should be directly related to the model state variables. For this reason, an observation operator,  $\mathbf{H}$ , is used to convert from model variables to measured variables for the case of SLA. We compute sea level with the full diagnostic surface pressure formulation described by Pinardi et al. (1995) but we leave a simplified  $\mathbf{H}$  operator in the definition of  $K^{\text{ROOI}}$ , as explained in Appendix C. In particular,  $\mathbf{H}$  contains only the geostrophic part of the signal and it assumes that bottom topography-related processes do not contribute to the signal. It is equivalent to the computation of dynamic height with respect to a deep reference level with the addition of the barotropic mode from the model. Thus the operator  $\mathbf{S}\mathbf{H}^T$  projects the SLA into the temperature, salinity and stream function modes that are consistent with the geostrophic constraint for the SLA.

In the MFSPP assimilation system we used two different sets of EOFs, one for the XBT and the other for the SLA assimilation. This is done in order to strengthen the formal requirement that the observation operator projection on the null space of the Kalman matrix (Appendix C) is marginal and that the vertical EOFs should have a relatively high signature

**Table 1.** Variance fraction (in %) explained by the first three bi-variate EOFs in three different areas of the Mediterranean Sea for the spring season (AMJ)

Area	EOF-1	EOF-2	EOF-3	Total
Algerian Basin	37.9	23.4	11.0	72.3
Tyrrhenian Sea	42.1	21.8	10.0	73.9
Ionian Sea	34.5	27.2	16.5	78.2

on the observations. We believe that the same set of EOFs cannot fulfill the null space conditions for the both SLA and XBT, since SLA contains information mainly about the  $T$ ,  $S$  variability below the mixed layer while XBT give the mixed layer temperature. We use two different sets of order reduction EOFs based upon our a priori knowledge of the information content of each observational data set.

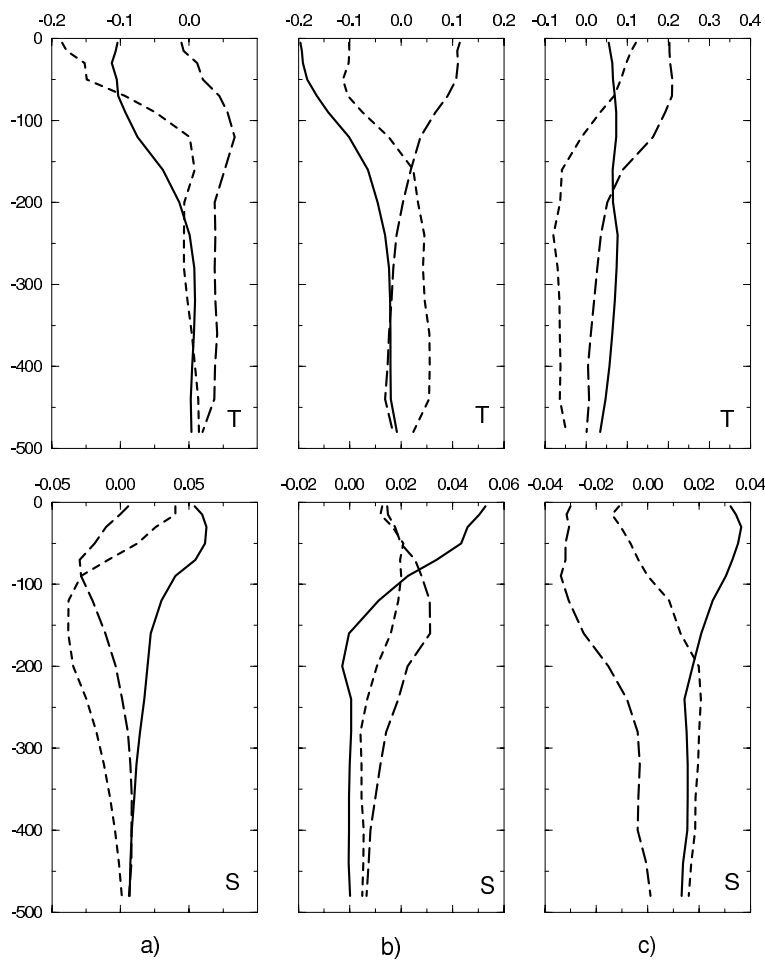
To assimilate SLA only one multivariate EOF for the whole basin was used, similarly to the scheme described in De Mey and Benkiran (2002). The SLA EOF is three-variate, i.e. computed from the covariance of temperature, salinity and stream function. The temperature and salinity components of the SLA EOF are shown in Fig. 2. The EOF extends only from 120 m downward. This choice of EOF for SLA is dictated from the past experience that showed that very few vertical modes can represent most of the dynamic height variability (Faucher et al., 2002) and that the SLA physically represents most of the geostrophic variability signal below the mixed layer. The choice of 120 m for the Mediterranean may be excessive but we took the conservative view of using EOFs that were already shown to be working in the region.

In order to assimilate the XBT data set, including the surface layer, a second set of bi-variate temperature and salinity EOFs was used. The EOFs were computed from an historical data set of the Mediterranean (Sparnocchia et al., 2003) in 9 different regions, chosen on the basis of the data coverage and known regional dynamical regimes. The EOFs retained only the variance around the seasonal signal and thus they vary every three months (winter consists of January, February and March, spring of April, May and June, summer of July, August, September and autumn of October, November and December). For each region, 10 dominant EOFs were considered. The transition between different regions was provided by a smooth change of the background error covariance matrix while the transition from one season to the next was sudden.

Examples of the first three EOF modes for three different regions – Algerian Basin, Tyrrhenian Sea and Ionian Sea are shown on Fig. 3. The variance fractions explained by the first 3 EOFs shown on Fig. 3 are given in Table 1. The dominant (first) EOF in Algerian Basin and Tyrrhenian Sea is quasi monotonic with maximum amplitude in the surface 100–150 m layer, which roughly corresponds to the layer of Modified Atlantic Water. The second and third salinity

**Table 2.** The variance of temperature ( $^{\circ}\text{C}$ ) and salinity (psu) at different depths for the 480 m surface layer in three different areas of the Mediterranean Sea: the Algerian Basin ( $\sigma_1^T, \sigma_1^S$ ); the Tyrrhenian Sea ( $\sigma_2^T, \sigma_2^S$ ); the Ionian Sea ( $\sigma_3^T, \sigma_3^S$ ). The depth of model levels in the first line of the table is given in meters

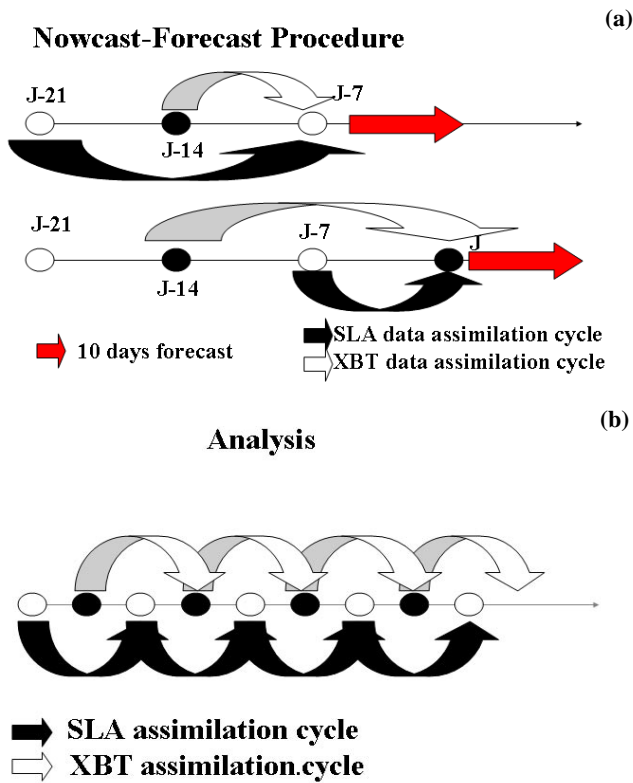
	5	15	30	50	70	90	120	160	200	240	280	320	360	400	440	480
$\sigma_1^T$	1.70	1.61	1.20	0.73	0.60	0.54	0.41	0.26	0.18	0.17	0.14	0.14	0.14	0.14	0.15	0.11
$\sigma_1^S$	0.26	0.29	0.32	0.32	0.30	0.26	0.18	0.11	0.09	0.08	0.07	0.06	0.05	0.05	0.04	0.04
$\sigma_2^T$	1.02	1.14	1.01	0.66	0.47	0.39	0.35	0.33	0.33	0.28	0.25	0.23	0.22	0.21	0.22	0.14
$\sigma_2^S$	0.27	0.22	0.20	0.19	0.16	0.14	0.11	0.11	0.10	0.08	0.06	0.05	0.04	0.03	0.03	0.03
$\sigma_3^T$	1.09	0.97	0.93	0.72	0.62	0.52	0.44	0.39	0.33	0.34	0.31	0.30	0.30	0.30	0.29	0.25
$\sigma_3^S$	0.22	0.20	0.20	0.19	0.15	0.14	0.13	0.11	0.08	0.08	0.07	0.07	0.07	0.07	0.08	0.08



**Fig. 3.** Bivariate Empirical Orthogonal Functions used in assimilation of XBT in (a) Algerian Basin, (b) Tyrrhenian Sea and (c) Ionian Sea. The continuous line shows the first EOF, the long dashed line the second EOF and the short dash line, the third EOF. The temperature EOFs are present in the upper panels and salinity EOFs in the bottom panels.

modes in these regions have maxima at intermediate depths between 100 and 300 m. The first bi-variate EOF in the Ionian sea, Fig. 3c, show relatively low vertical variability similar to the dominant EOFs in most other Eastern Mediter-

ranean regions (not shown here). The second and third modes in the Ionian Sea have high salinity amplitude at the surface, different from the subsurface salinity maxima present in the second and third modes of the Algerian Basin and Tyrrhenian



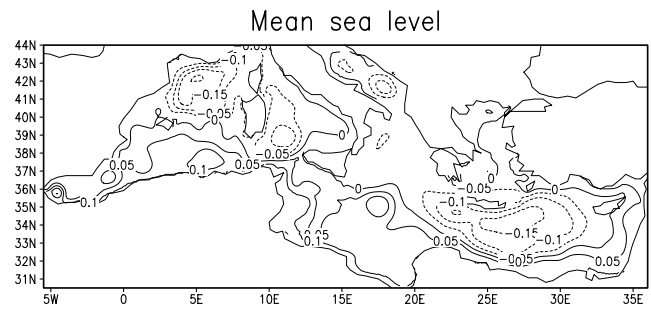
**Fig. 4.** The scheme of sequential assimilation of SLA and XBT data (a) in the nowcast-forecast procedure and (b) in the analysis.

Sea. When composing the **S** matrix, the EOFs are multiplied by the variance of the temperature and salinity at each level (see Table 2).

It is interesting to note that the first (dominant) EOFs in the Algerian and Tyrrhenian Sea have a structure below 120 m similar to that of the EOF used in SLA assimilation but this is not the case for the Ionian Sea. In fact, as explained by Sparnocchia et al. (2003), the EOFs are very different in the western and eastern Mediterranean basins, accounting for the different water mass variability in the two sub-basins.

The SLA and XBT sets of EOFs are used in a sequential procedure that ensures the combined assimilation of SLA and XBT with the different sets of EOFs (Fig. 4). The assimilation cycle is chosen to be a week, consistent with the XBT data distribution (Manzella et al., 2001) and the SLA data availability (Pinardi et al., 2003).

The first sequential procedure is called **nowcast-forecast** and it prepares each week the initial condition for the forecast (Fig. 4a). Every assimilation cycle, two sequential runs are made – the first with assimilation of one of the data sets in smoother mode and the second with the other data set assimilated in filter mode. The multivariate EOFs for  $\psi$ ,  $T$ ,  $S$  are used for assimilation of SLA and the bi-variate  $T$ ,  $S$  EOFs are used for assimilation of XBT profiles. This way each data set is projected into its optimal vertical modes, and contributes to the estimate of the nowcast. The SLA assimilation is applied only in regions deeper than 1000 m. This assumption is based upon the surface dynamic height com-



**Fig. 5.** Mean climatology of the sea surface height in meters, computed from 1993–1997 model simulations.

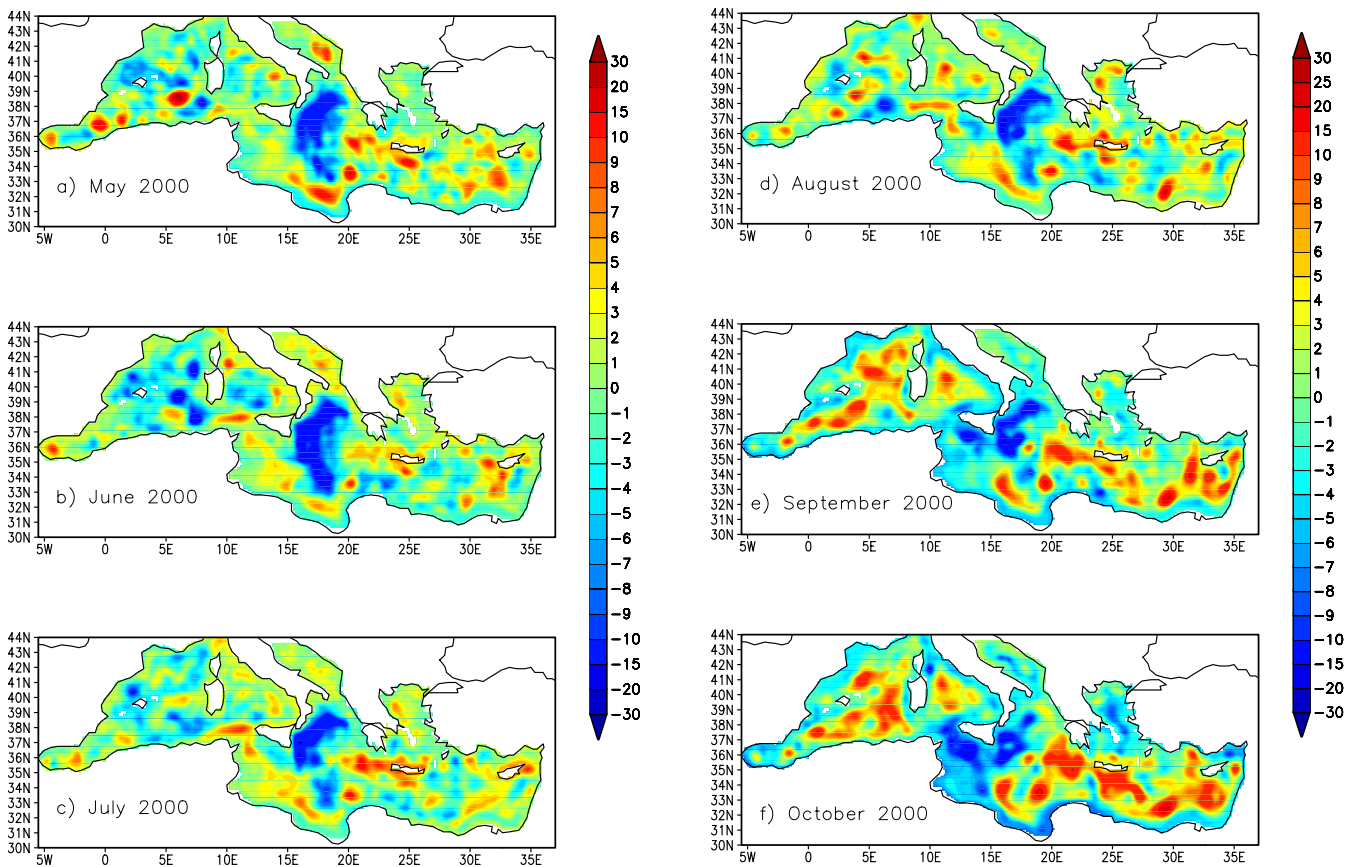
putation studies of Özsoy et al. (1993) who showed that a deep reference level (zero motion assumption) would generate more realistic surface currents and sea level variability. The model SLA is then calculated by subtracting the model mean sea surface level that is shown in Fig. 5. This field was computed from a model simulation of the Mediterranean circulation from 1993 to 1997 to be consistent with the mean subtracted from the observed SLA.

The second procedure is called **analysis** and it is presented in Fig. 4b. It uses only the results of the smoother assimilation mode for both SLA and XBT. This means that we produce an optimal estimate of the circulation once every week with the usage of both past and future observations. Between one analysis and the other we run the model with surface meteorological analyses, thus producing the best dynamical extrapolation between successive optimal ocean state estimates. The daily data set formed by this analysis/simulation is called, collectively, the analysis data set, even if the ocean data are inserted only once a week. In both procedures of analysis and nowcast-forecast, the model surface fluxes are not only calculated from the analysis atmospheric fields but are also corrected with the observed weekly mean SST using Eq. (B1).

#### 4 Forecast skill scores

Our study period goes from 4 April 2000 to 31 October 2000. It coincides with the first period of operational forecasting in MFSP and it was chosen because it has the largest data density for both in situ and satellite data. In the following the model forecast is checked against observations and analyses following the Murphy (1993) definition of indices for “good” forecast. These are:

1. **Consistency:** it checks the qualitative consistency between the analyses and the observations based upon the experience of the forecasters. This is not an objective index but it is a check based upon the previous experience of the forecaster on the dynamics of the region and the knowledge of the relevant dynamical processes.



**Fig. 6.** Monthly mean of MFSPP SLA analysis in cm for (a) May 2000, (b) June 2000, (c) July 2000, (d) August 2000, (e) September 2000, (f) October 2000.

2. *Quality*: the objective correspondence between analyses and forecasts using statistical indices to quantify the level of discrepancy between forecast and analyses.
3. *Accuracy*: the objective correspondence between observations and analysis using statistical indices to quantify the analysis error with respect to the data.
4. *Value*: the benefits realized by the users of the forecast. The value of the Mediterranean forecasts cannot be yet quantified since the user basis of the forecasts is still very limited. The other indices will be discussed in the sections below.

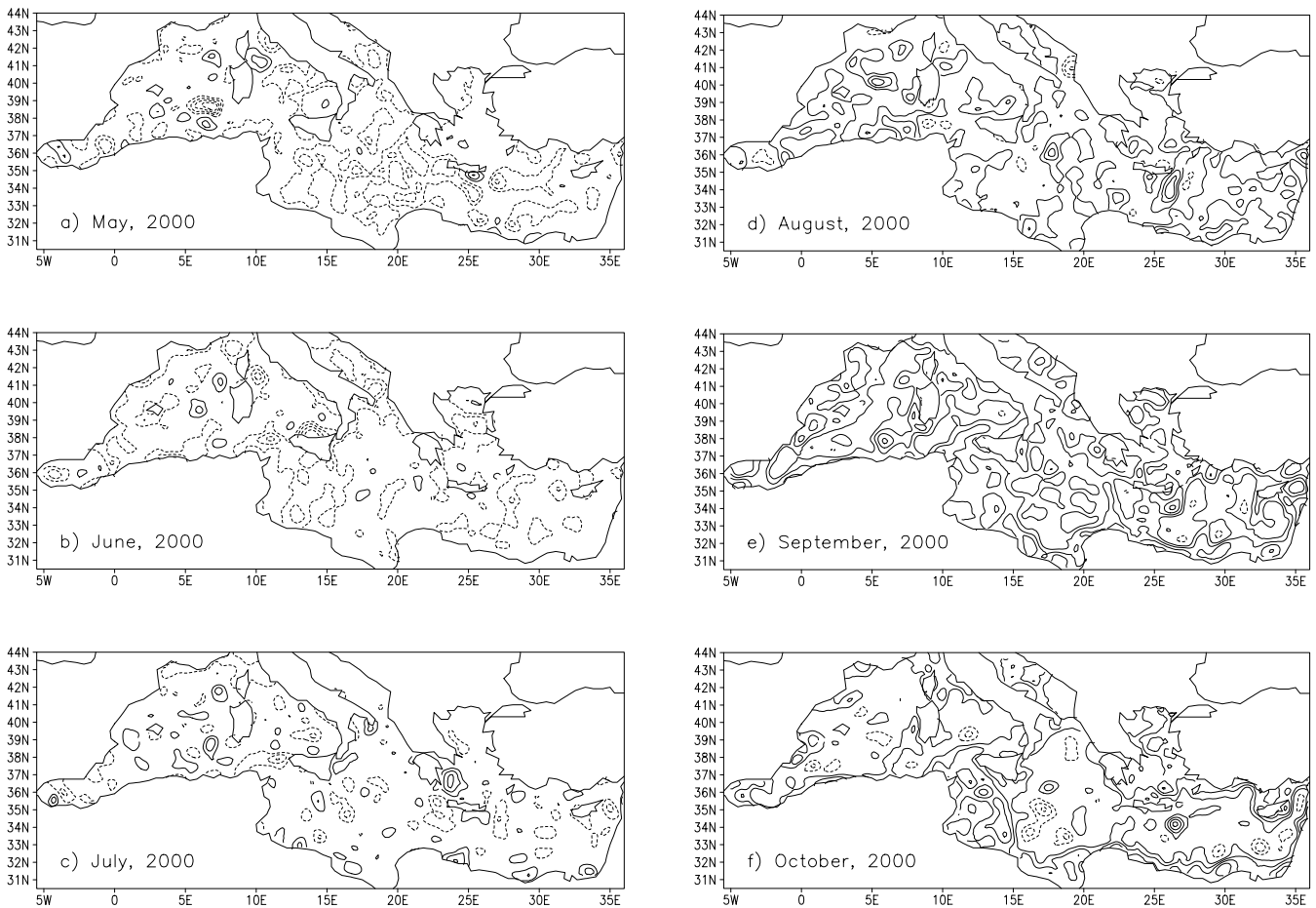
#### 4.1 Consistency

Our consistency check is defined on the basis of the comparison between satellite data and analysis for SLA. Instead of using along-track values of SLA from the model and observations, we will compare the analyses done with two different analysis systems. The first is based only on the statistical knowledge of the satellite data structure and is explained by Le Traon et al. (1998). The observed SLA is mapped by objective analysis (OA) techniques using three weeks data to estimate a field every week. The weekly OA field is then av-

eraged to produce monthly mean distributions of SLA for the period May–October 2000. The second field is produced by the MFSPP analysis system of Fig. 4b, and its monthly mean distributions for the same period of time are shown in Fig. 6. The correspondence between MFSPP SLA analysis and OA SLA data is evaluated by the differences of corresponding fields (Fig. 7).

Both data and model show relatively strong cyclonic circulation (or low, negative values of SLA) all over the basin during the May–June period. Anticyclonic eddies and gyres are locally intensified in the Algerian Basin, Southern Ionian, Peloponnesus and south-east of Crete, the so-called Iera-Petra gyre area, and the Shikmona gyre area (south of Cyprus). These features are quasi-permanent and they are observed with variable intensity during the whole period. The large cyclonic anomaly in the Ionian Sea is very interesting; this is maintained both in the model and data for the whole period (Figs. 6 and 7). Anticyclonic anomalies are stronger in the model than in the data during the May–June period (Figs. 7a and b) since the difference fields are mainly negative.

Both satellite data and analyzed fields show a strong tendency to transit from a dominant cyclonic circulation in June toward an positive SLA in September–October (anticyclonic



**Fig. 7.** Difference between monthly mean OA SLA computed only from satellite observations (see the text) and monthly mean of MFSP SLA analysis for (a) May 2000, (b) June 2000, (c) July 2000, (d) August 2000, (e) September 2000, (f) October 2000. The contour interval is 5 cm and the 0 isoline is not plotted.

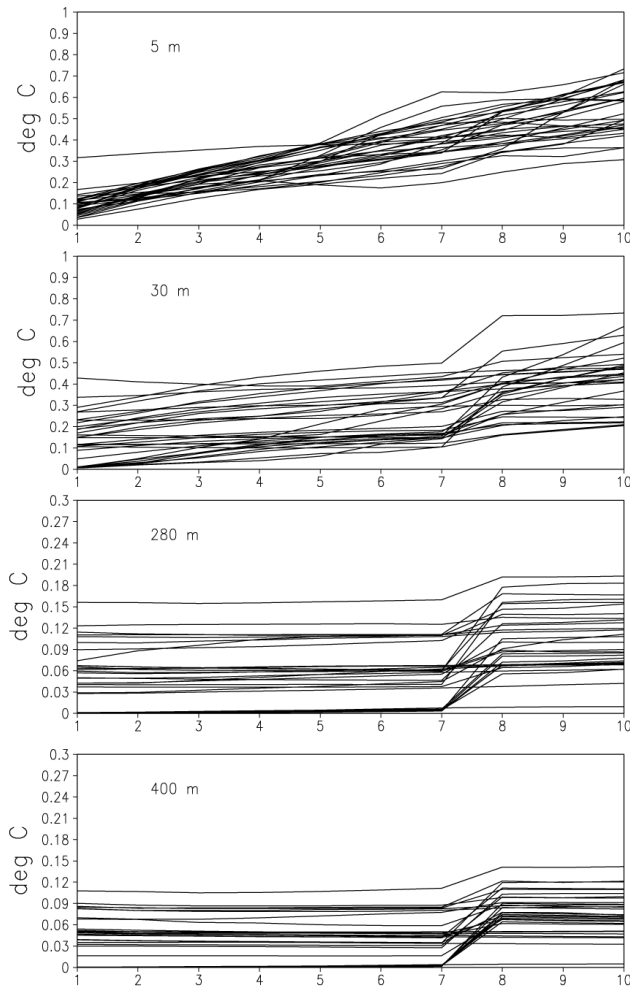
intensification). The strongest change in the observed data, which, hereafter, will be referred to as the summer transition of SLA, appears in July and August. The observed SLA values reach the highest positive values during September 2000.

The MFSP analyses are not capturing this summer SLA transition, which can be seen by the positive high values of the difference between data and model analysis in July and August (Figs. 7d and e). There are two major differences between the satellite data and the MFSP SLA analysis in this period. Firstly the change in the SLA analysis field is smoother with a gradual increase of the positive values. The highest values of SLA in the model solution are observed in October 2000 (Fig. 6). Secondly, the anticyclonic intensification in the analysis solution appears only in the deep areas like Algerian – Provencal basin, Ionian Sea and the Levantine Basin. In the shallow areas of the Strait of Sicily, the Aegean and the Adriatic Sea the values of model SLA remain relatively low during the whole period after the summer transition (Figs. 7d–f). As we mentioned in the Sect. 4, the SLA assimilation is done only in the deeper parts of the basin (deeper than 1000 m). The data assimilation has an im-

portant impact on the model solution there, which even with some delay develops circulation structures equivalent to that in the observations. In the shallow areas however, where no SLA data assimilation is done, the summer transition is not captured at all by the model dynamics.

In summary, this first comparison shows that the analyses are relatively consistent with the observations but that coastal areas are not greatly affected by open ocean assimilation and that rapid transitions are difficult to capture at least with a weekly assimilation cycle. A future improvement of the consistency between analyses and observations could come from considering the definition of the observational operator also for the coastal areas and a shorter assimilation cycle for SLA in order to resolve rapid one-month transitions. Another valuable improvement would be to have EOFs that correct for SLA up to 50 m, a more realistic estimate of the Mediterranean mixed layer depth for the summer.





**Fig. 8.** Root mean square temperature forecast error at (a) 5 m; (b) 30 m; (c) 280 m; and (d) 400 m. Different curves correspond to 31 different 10-day forecasts carried out from April to October 2000.

#### 4.2 Quality

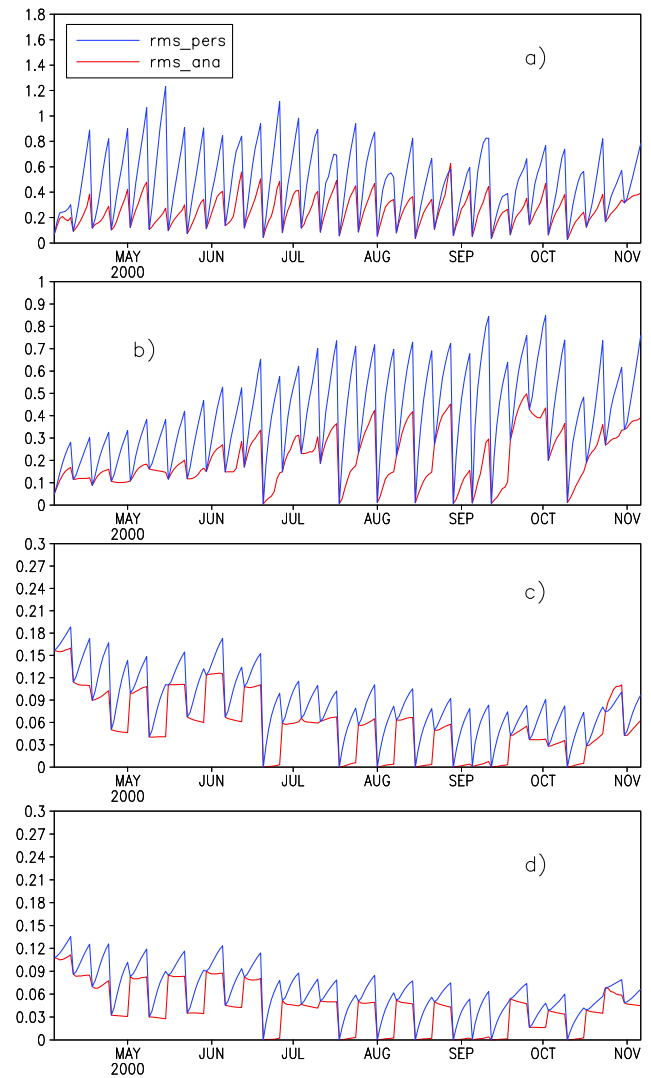
The rms error between two quantities,  $\phi_f$  and  $\phi_r$  in general is defined as:

$$\text{rms}(\phi) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\phi_f - \phi_r)^2} \quad (1)$$

where  $N$  is the number of data used in the evaluation. Here we will use  $\phi_f$  for the forecast fields and  $\phi_r$  for the analysis or observations.

Figure 8 shows the root mean square (rms) error between temperature forecast and analysis for 31, ten days forecasts carried out between 4 April and 31 October 2000. The parameter plotted on Fig. 8 will be referred hereafter as forecast rms temperature error.

The forecast rms temperature error in the surface layer (5 m) increases almost linearly with time. This is mainly due to errors in the atmospheric forcing computed with ECMWF forecast surface fields. We remind that the analysis is done

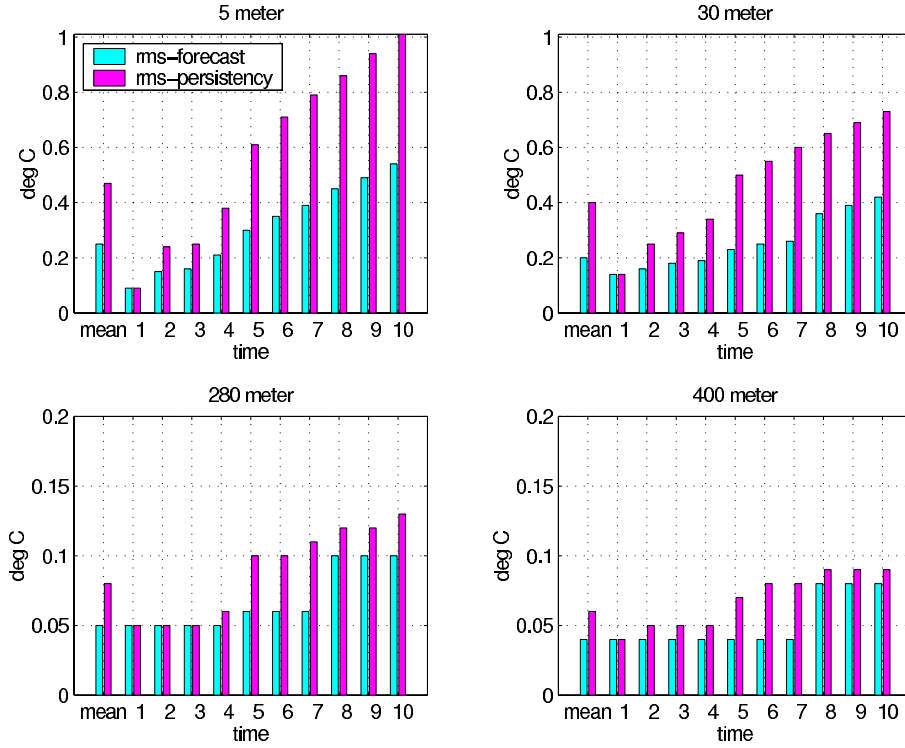


**Fig. 9.** Root mean square temperature forecast (red line) and persistence (blue line) errors at (a) 5 m; (b) 30 m; (c) 280 m and (d) 400 m for the 31, 10-day forecasts carried out from April to October 2000.

not only with a more accurate representation of surface forcing, calculated using the atmospheric analysis fields, but it uses the heat flux correction with observed SST. The error in the estimation of the surface heat flux during the forecast influences the thermal structure in the whole surface mixed layer in a coherent way and in fact, at 30 m depth, the forecast rms temperature error behaves almost linearly, as at the surface.

Below the surface layer, the forecast rms temperature error grows relevantly only at the analysis time, i.e. at the end of every assimilation cycle (or every 7th day). In Fig. 8 we see in fact that at depth the forecast rms temperature error increases relevantly only at day 7 when new data are inserted at depth.

The maximum forecast rms temperature error range, from April to October 2000, is, at the surface, between 0.2°C–



**Fig. 10.** Average root mean square temperature forecast (blue bars) and persistence (red bars) errors at (a) 5 m; (b) 30 m; (c) 280 m and (d) 400 m. The average is carried out over the 31 forecasts done from April to October 2000.

0.7°C, and below 200 m, between 0.03°C–0.18°C. The same forecast rms temperature error is shown on Fig. 9 together with the rms error between forecast and initial condition. Here  $\phi_r$  is taken to be the nowcast for each week; the skill score computed this way will be referred to hereafter as persistence rms temperature error. With very few exceptions, the forecast rms temperature error is less than the persistence rms temperature error during the whole forecasting period and at every model depth. In the surface layer, where the time variability is the highest, the persistence rms temperature error during the spring and summer can reach values close to 1.2°C, i.e. almost twice the maximum forecast rms temperature error at this level.

The rms forecast and persistence errors are minimal at the beginning of every assimilation cycle with values which depend on the differences in the initial conditions of forecast and analysis. The latter are computed with data assimilation in the smoother mode while the forecast uses the nowcast produced with a combination of smoother and filter schemes. During the assimilation cycle, the rms persistence error increases due to the dynamical evolution of the model fields. At the surface, the rms persistence error is relatively high during the period of strong heating in May and June, while at 30 m the highest rms persistence error is present during July–August, i.e. after the formation of the seasonal thermocline. The rms forecast error in the surface layers is mainly due to the uncertainty in the surface forcing but it always remains below the rms persistence error.

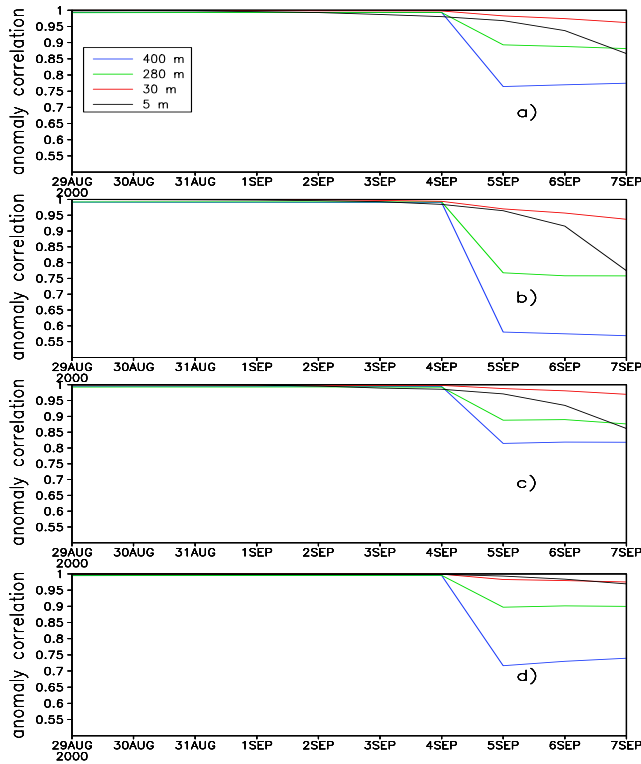
At 280 and 400 m (Figs. 9c and d) the changes in the rms forecast and persistence errors are relatively high at the be-

ginning of every assimilation cycle. During the rest of the time their variability is relatively small. The forecast and persistence rms errors, after April, reveal a bi-weekly variability which is related to the combined assimilation cycle of SLA and XBT data. After the beginning of July, the bi-weekly variability of the forecast and persistence rms errors become different. This is due to the fact that the regular XBT data collection stopped and the main assimilated data set was composed of SLA observations. The analysis and the nowcasts were produced with a relatively small amount of XBT data or just as model simulations, since the XBT data were not available.

The mean value of forecast and persistence rms temperature error for the period April–October 2000 is shown in Fig. 10 for different model depths. The forecast rms temperature error is highest at the surface, where it reaches 0.55°C at the last forecast day. Its mean value decreases with the depth. Its mean is about 0.2°C at 30 m and less than 0.05°C at 280 and 400 m. The maximum of the persistence rms temperature error in the surface layer is about 1°C while at 30 m it is about 0.68°C. The mean value of the persistence rms temperature error has values of 0.45°C at the surface and 0.38°C at 30 m. At all depths it is higher than the forecast rms error.

Another statistical parameter, which characterizes the forecast quality is the anomaly correlation, which is defined as follows:

$$C_a = \frac{\sum_{n=1}^N (T_n^a - T_n^{ca})(T_n^f - T_n^{cf})}{\sqrt{\sum_{n=1}^N (T_n^a - T_n^{ca})^2} \sqrt{\sum_{n=1}^N (T_n^f - T_n^{cf})^2}}, \quad (2)$$



**Fig. 11.** Anomaly correlation (Eq. 2) of forecast and analysis for (a) the whole Mediterranean Sea; (b) the Algerian Basin; (c) the Tyrrhenian Basin; (d) the Ionian Sea. The black curves correspond to the 5 m depth, the red to 30 m, the green to 280 m and the blue to 400 m.

where  $T^a$  and  $T^f$  are the analysis and forecast temperature fields, respectively,  $T^{ca}$  is the climatology computed from the analysis,  $T^{cf}$  is the forecast climatology and  $N$  is the number of the model grid points for each region or the whole basin.

The temperature anomaly correlation is shown for the whole basin in Fig. 11a, for the Algerian Basin in Fig. 11b, for the Tyrrhenian Sea in Fig. 11c and for the Ionian Sea in Fig. 11d. The anomaly correlation in the surface layer remains relatively high during the first 7 days. However, due to the new observations inserted in the analysis at day 7, the correlation falls down to 0.8–0.6 in the last three days of the forecast period. Our choice of climatology makes the terms in Eq. (2) insensitive to the changes occurring during the first seven days of forecast. In the future, a more sensitive skill score index for correlation should be used.

### 4.3 Accuracy

The forecast and persistence rms temperature errors and the anomaly correlations are measures of the departures of forecast from the analyses, i.e. the best estimate of the ocean state computed by using model and data. A different way to evaluate the quality of the forecast is to compare the model solution directly to the observations. Since the amount of independent data (not used in the analysis) for the period con-

sidered is limited, we use here the assimilated data itself before they are inserted in the model. Thus we compare the model simulation done with an analysis initial condition and atmospheric fields to the observations.

During the assimilation procedures, a misfit is calculated, that is the difference between model simulation and the observations at the time and locations of the observations. The model simulation is given at the observation location using a simple linear interpolation scheme. The rms statistics are then calculated on such misfits. We define the normalized mean square (nms) as:

$$\text{nms} = \frac{\sum_{i=1}^N (\phi_m - \phi_o)^2}{\sum_{i=1}^N \phi_o^2}, \quad (3)$$

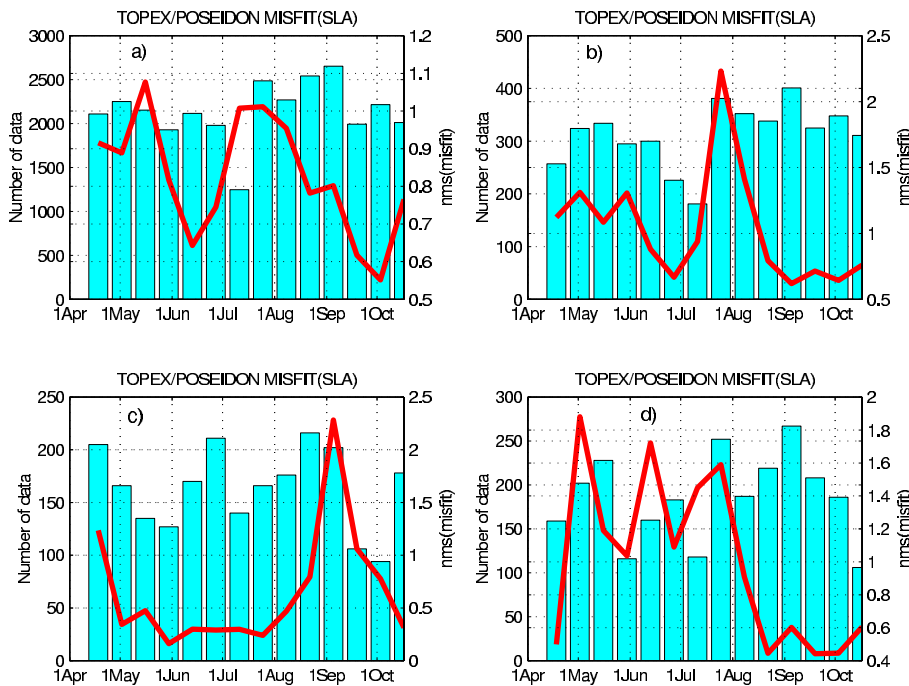
where  $\phi_m$  is the model field,  $\phi_o$  the observation and  $N$ , the number of observations used in the weekly analysis scheme.

In Figs. 12 and 13 the nms for SLA is shown for Topex/Poseidon and ERS-2 data, respectively. Both figures show a decrease of nms from April to June/July in the whole basin and the Algerian and Tyrrhenian. However, nms grows back again in August for the Algerian Basin and in September for the Tyrrhenian Sea. In addition, in the Ionian, the nms remains relatively high up to August–September. This is the problem of the missed summer SLA transition period discussed above. It is interesting to note that, during the period May–August (Fig. 6) when the SLA show the presence of anticyclonic eddies in the Ionian, the nms is the highest. In contrast, during September–October, the strong intensification of the quasi-permanent anticyclonic gyres such as the Iera-Petra, the Mersa-Matruh and the gyres in the Shikmona area is relatively well represented also by the model solution and the nms decreases. In the Tyrrhenian Sea, it cannot reproduce the anticyclonic intensification occurring in September.

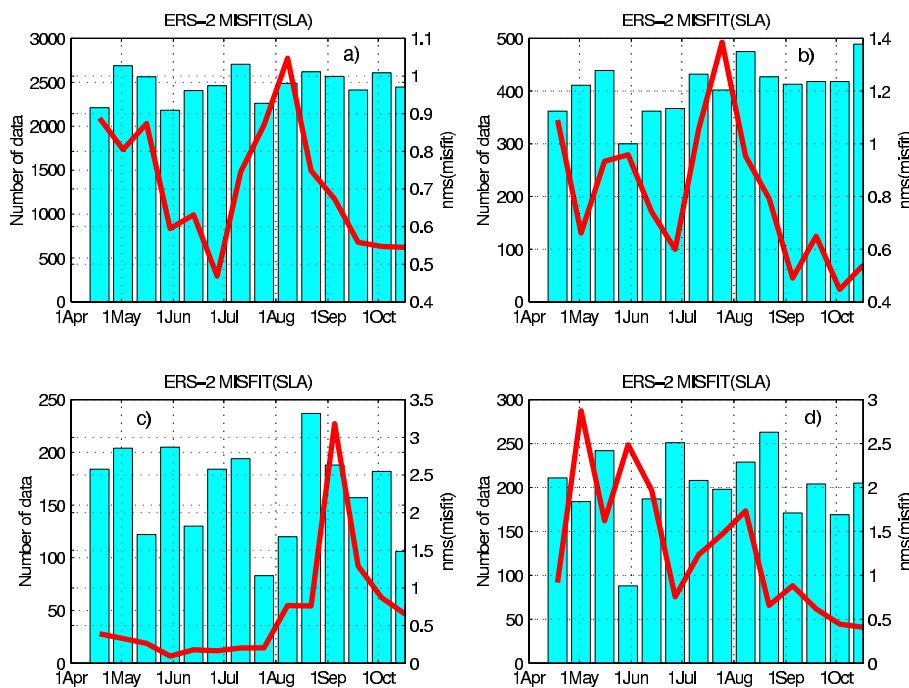
In summary, the nms of the SLA misfit shows that the model can capture, with different skills, different parts of the ocean variability. It shows problems for capturing the intense mesoscale activity, presumably due to the coarse resolution of the model which is just eddy permitting but not resolving. Different regions have different rates of decrease of the nms misfit but they show a growth of errors in the period of anticyclonic intensification. Another important point is that the value of the nms is different between the two data sets, hinting that T/P is closer to the model solution than ERS-2.

The rms temperature misfit for the XBT data, calculated with Eq. (1) is shown in Figs. 14 and 15 for two different model levels, 5 and 360 m, respectively. The zero values correspond to the periods of no XBT data. In the surface layer of the whole basin (Fig. 14a) the misfit varies between 0.4°C and 0.7°C. The error increases during the end of spring and beginning of summer. The maximum errors is about 0.7°C in the Algerian Basin (Fig. 14b) for the first week of June, about 0.55°C in the Tyrrhenian Sea (Fig. 14c) for the first week of May and about 0.7°C in the Ionian Sea (Fig. 14d) for the first week of July.

At intermediate depth, the rms temperature misfit error maxima is about 0.5°C (Fig. 15a) in the last week of May.



**Fig. 12.** Normalized root mean square misfit of SLA (Eq. 3) computed for Topex – Poseidon data for (a) whole basin; (b) Algerian Basin; (c) Tyrrhenian Sea; (d) Ionian Sea. The blue bars indicate the number of SLA observations available during every bi-weekly analysis cycle.



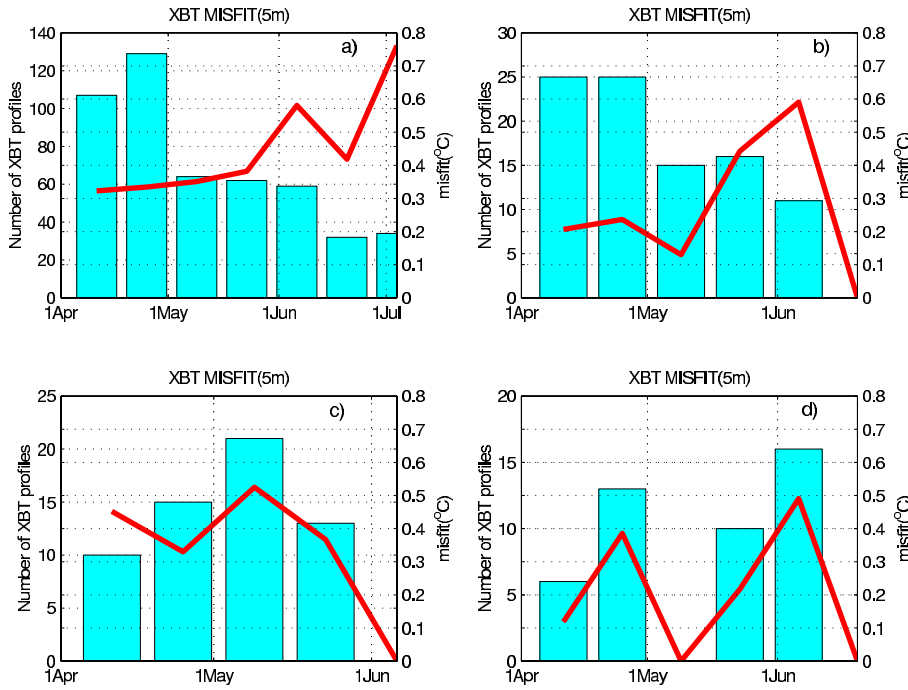
**Fig. 13.** Normalized root mean square misfit of SLA (Eq. 3) computed for ERS-2 data for (a) whole basin; (b) Algerian Basin; (c) Tyrrhenian Sea; (d) Ionian Sea. The blue bars indicate the number of SLA observations available during every bi-weekly analysis cycle.

The variability of the error is relatively low in the intermediate layers of the Algerian Basin (Fig. 15b) and the Tyrrhenian Sea (Fig. 15c), where it changes between 0.2°C and 0.3°C. In the Ionian Sea the rms temperature misfit error in the middle of April is about 0.9°C and decreases to 0.3°C.

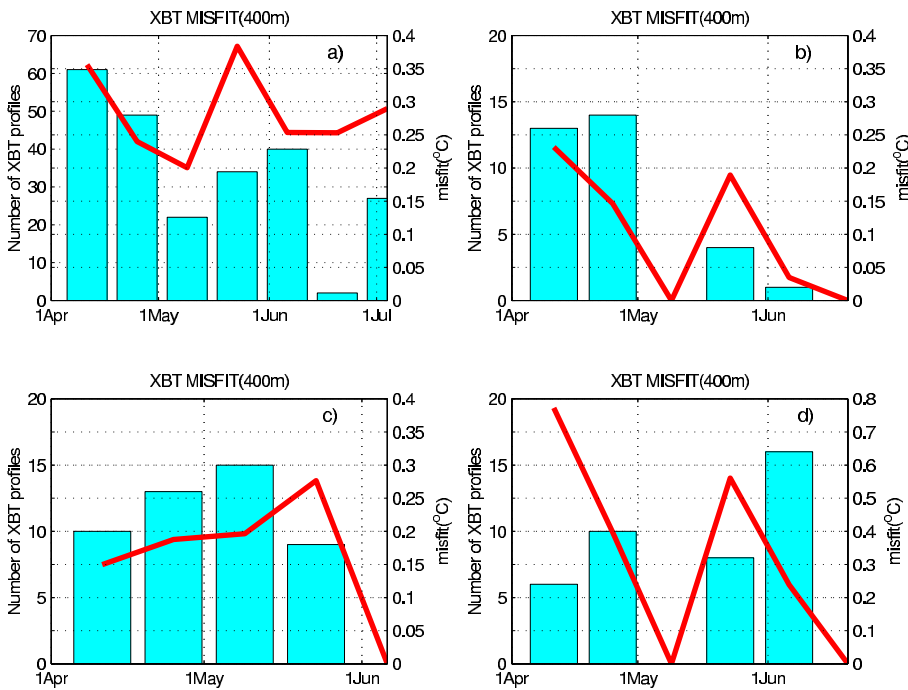
From the rms temperature misfit error it is possible to deduce that, with two weeks repeat sampling, the error has time to approximately double or triple but is still kept on reason-

able values, below 1°C. Toward the end of the XBT sampling experiment, in July, the 5 m rms temperature misfit error increases due to the decreasing number of observations. Below the threshold of 50–60 XBT every two weeks, the error starts to double in most of the regions (Figs. 14a, b and d).

In a separate study of the XBT data impact on the analysis, Raicich and Rampazzo (2003) show that the assimilation of two weeks repeat XBT tracks is not sufficient to keep the



**Fig. 14.** Normalized root mean square misfit between forecast and XBT observations at 5 m depth for (a) whole basin; (b) Algerian Basin; (c) Tyrrhenian Sea; (d) Ionian Sea. The blue bars indicate the number of XBT profiles available during every bi-weekly analysis cycle.



**Fig. 15.** Normalized root mean square difference between forecast and XBT observations at 400 m depth for (a) whole basin; (b) Algerian Basin; (c) Tyrrhenian Sea; (d) Ionian Sea. The blue bars indicate the number of XBT profiles available during every bi-weekly analysis cycle.

solution close to the observations. These authors show that the time period during which the assimilated information is retained by the model is shorter than two weeks. In addition, the data used in Raichich and Rampazzo (2003) are synthetic XBT profiles regularly distributed in time while the XBT observations, used in operational forecasts, were rather sparse in both time and space due to the irregularity of the realistic sampling scheme and the loss of data by the satellite commu-

nication system. Thus the highly variable number and position of the XBT data used in forecast initialization, as shown on Fig. 14, reduces their impact on the model solution and increases the model errors in accordance with the Raichich and Rampazzo (2003) results.

The temperature misfit error during some weeks of the April to June period of MFSP TOP (Fig. 14) tends to double or triple but it is kept below 1°C. At the same time the

number of data decrease at approximately the same rate. In the surface layer, the misfit rms error is close to the maxima of rms temperature forecast error (Fig. 8a). As mentioned above, the rms forecast error increases almost linearly during each ten day forecast indicating a strong impact from the uncertainty in the surface fluxes. The seasonality in the misfit distribution is presumably related to model deficiencies in representation of heat penetration and lateral transport in the upper layer during the end of spring and beginning of summer, when the surface heating is relatively strong. We have to mention that our vertical mixing parameterization does not include the contribution of some important processes like wind mixing. The heat flux correction is computed with the weekly mean SST. At the same time, the fast changes in the upper thermal structure of the sea during this period have a significant daily and day to day component which decreases the impact of the satellite SST corrections.

The maxima of temperature misfit at intermediate depths is about  $0.6^{\circ}\text{C}$  for the whole basin (see Fig. 15) which is higher than the value (about  $0.12^{\circ}\text{C}$ ) of rms forecast error (Fig. 8). This indicates that the data assimilation is only partly capable of correcting the differences between the observed in situ values and model parameters. This may be due to several reasons, the first being the possible inconsistency between SLA and XBT assimilation schemes at depth.

## 5 Conclusions

In this paper we presented the methodology of data assimilation developed for the MFSPP operational forecasting exercise lasting from April to October 2000. The basic assimilation scheme is SOFA that is an OI multivariate reduced order technique that uses vertical EOFs to reduce the order of the assimilation problem.

The method successfully assimilates a combination of SLA and XBT profiles and uses weekly SST to correct for heat fluxes during analysis cycles. For the first time, a multivariate OI scheme has been used to assimilate SLA and XBT observations that form the basis of the NRT ocean monitoring programs in the world's ocean's and now in the Mediterranean Sea. The OI is used with different sets of EOF in order to optimize the process of information extraction from these two data sets.

The vertical multivariate EOF used for SLA assimilation corrects the subsurface temperature, salinity and barotropic stream function fields. A single EOF is used for SLA assimilation since it is known that most of the variability in dynamic height can be represented by very few vertical modes. The correction from SLA is done only below 100 m since, again, it is believed that the SLA signal is indicative of geostrophic dynamics below the mixed layer.

XBT are assimilated with a multivariate OI scheme. For this we use 13 different regions, 4 different seasons and 10 vertical bivariate EOF. This allows us to get the benefit of XBT assimilation throughout the first 500 m of the water column where most of the XBT are collected. Such a large

number of vertical modes, as demonstrated by Sparnocchia et al. (2003), also captures the mixed layer information.

The OI scheme is used both in smooth and filter mode and the analysis-forecast procedures are shown in Fig. 4. To our knowledge this is the first multivariate operational assimilation of SLA and XBT in the world's ocean's. The operational nowcasting-forecasting and analysis procedures are done intermittently with an assimilation cycle of one week. Forecasts are launched once every week and they use atmospheric surface parameters to force the ocean forecasts for ten days.

Finally the consistency, quality and accuracy of the forecast has been evaluated with respect to observations and analyses.

The forecast rms temperature error versus persistence rms temperature error shows that for a six months period, from April to October 2000, forecast always beats persistence. The forecast rms temperature error is maximum at the surface, reaching  $0.6^{\circ}\text{C}$  after ten days and lower in the subsurface, up to  $0.3^{\circ}\text{C}$  after the same ten days. The anomaly correlation drops to about 0.75–0.8 after ten days, only in the subsurface. We are tempted to deduce that the predictability time of the large-scale temperature field is longer than ten days at all levels in the Mediterranean, given the observing system network implemented during MFSPP. Coastal areas however, remain still with high consistency errors between analyses and observations and in the future the forecasting system should develop the scheme to use multivariate assimilation of SLA also in these areas.

The rms error of the SLA misfit show that the model can sometimes reproduce the mesoscale, some other times the larger scale subbasin scale gyre variability. However, the assimilation system proposed here shows problems to accurately retain information from the SLA during rapid transition periods, such as the July–August period in year 2000. More detailed regional analysis (not shown here) showed that in different regions this summer SLA transition lasted only 3–4 weeks. Under such conditions, the weekly assimilation of SLA is not efficient enough to “keep” the model close to the data. A possible way to improve the model forecast is to use smaller data assimilation cycles with more frequent insertion of data.

The future forecasting system for the Mediterranean Sea should consist of all three main components described here. While satellite data will be continuously available, the VOS-XBT may be more sporadic in the future but nevertheless still necessary to correct the subsurface temperature in a realistic way. Thus recommendations are that subsurface data acquisition programs will be sustained, perhaps with the help of more advanced technologies, such as ARGO subsurface floats and advanced multiparametric VOS system. On the other hand, the data assimilation scheme should be improved in order to accommodate the model error covariance matrix that allows us to consider assimilation of SLA into shallower areas.

*Acknowledgements.* The work has been supported by the EU-MAST Contract MAS3-CT98-0171, “Mediterranean Forecasting

System Pilot Project”.

The Editor-in-Chief thanks a referee for his help in evaluating this paper.

## Appendix A Near real time data collection and quality control

The data collected weekly for the MFSPP forecasting activities include NRT observations used in the data assimilation procedure and meteorological ECMWF analysis and forecast surface fields used in the computation of the surface forcing for the OGCM. The NRT observations consist of: (a) satellite data for sea surface temperature (SST) and sea level anomaly (SLA); (b) vertical temperature profiles by Expandable BathyTermograph (XBT) collected along the 7 VOS routes described by Manzella et al. (2001). The meteorological parameters used in the MFSPP forecast are: mean sea level pressure, total cloud cover, zonal and meridional wind components at 10 m, temperature at 2 m and dew point temperature at 2 m. In this section we describe briefly the data collection and preprocessing procedures for the different data sets.

The SST data are collected in the Centre de Meteorologie Spatiale (CMS) of Meteo France, Toulouse and the Istituto di Fisica dell' Atmosfera of the CNR in Rome. The data are obtained from the night orbits of the NOAA-AVHRR-14 and the NOAA-AVHRR-15 satellite sensors. The final SST data set is a weekly mean (centered on every Monday) and consists of data interpolated on the Mediterranean Sea OGCM grid (with resolution  $\frac{1}{8}^{\circ} \times \frac{1}{8}^{\circ}$ ) using an objective analysis method (see Buongiorno-Nardelli et al., 2003).

The along track Topex-Poseidon (T/P) and ERS - 2 satellite data for SLA are collected and analyzed at the Collection and Localization Satellitaire (CLS) located in Toulouse, France. The data are corrected first for the orbit error, computed by using a local inverse method (Le Traon and Ogor, 1998). Then SLA values are computed by subtracting a 5-year mean of T/P and ERS - 2 data from 1993 to 1997. The along - track correlated errors are corrected by application of a local adjustment method using simultaneous T/P and ERS2 data over a period J0-22 to J0-2 days where J0 is considered to be  $J + 1$  on Fig. 1. This local adjustment diminishes the residual error due to orbit and inverse barometer effects which are subtracted from the sea level signal. Smoothing cubic splines are then used to estimate a bias for each point along the track and to produce corrected along-track SLA observations.

The XBT data are collected along seven tracks with an along-track spatial nominal resolution of 12 nm (Manzella et al., 2001). Each track was repeated once per month from September 1999 until December 1999 and twice per month from January until June 2000 (with the exception of the track crossing longitudinally all the basin that was made only once a month). The temperature profiles have a vertical resolution of 0.6 m and reach the maximum depth of 460 m or 760 m (depending upon T6 or T7 probes being used). Decimated

data are received in NRT and consist of temperature observations at 15 vertical levels. They are transmitted on the Global Teleconnection System (GTS) and also collected at ENEA – La Spezia, Italy, where a first quality control is made before the data are provided on a free ftp site. At the forecasting center, and before the data are used in the assimilation procedure, a quality control check is carried out on each XBT profile through a graphical visualization of the vertical profile and a check on the position.

The ECMWF meteorological data are provided by Meteo France, Toulouse, France, for the basin scale forecast system with 12 h delay, i.e. on Wednesday of each week. Before starting the forecast procedure a quality control of the data is done by a graphical visualization of the fields.

## Appendix B The model

The model used is based upon the Modular Ocean Model (MOM), adapted to the Mediterranean Sea by Roussenov et al. (1995) and Korres et al. (2000). The model grid has 31 vertical levels, and an horizontal resolution of  $\frac{1}{8}^{\circ} \times \frac{1}{8}^{\circ}$ . Horizontal turbulent mixing is biharmonic with tracer coefficients equal to  $1.5 \times 10^{10} \text{ m}^4 \text{ s}^{-1}$  and momentum coefficients equal to  $5 \times 10^9 \text{ m}^4 \text{ s}^{-1}$ . Vertical turbulent coefficients are constant and equal to  $0.3 \cdot 10^{-4} \text{ m}^2 \text{ s}^{-1}$  for tracers and  $1.5 \cdot 10^{-4} \text{ m}^2 \text{ s}^{-1}$  for momentum. A convective adjustment procedure (Cox, 1984) is applied in statically unstable areas of the water column with 10 repeat cycles maximum each time step. This choice of physical mixing was elaborated in the articles cited above. Even if simple, it is capable of reproducing the bulk of water mass variability when it is associated with high frequency atmospheric forcing (Castellari et al., 2000). The transport through the Strait of Gibraltar is parameterized by extending the model area westward of Gibraltar to a longitude  $9.25^{\circ} \text{ W}$ . In this model area, which is a part of the North Atlantic, between latitudes  $33^{\circ}30' \text{ N} \leq \phi \leq 37^{\circ} \text{ N}$ , the surface forcing is switched off and temperature and salinity are relaxed toward annual mean climatological fields.

The model is rigid lid but a diagnostic computation of sea level is done at each time step following Pinardi et al. (1995). The sea level is proportional to the surface pressure on the rigid lid due to large scale dynamical response of the sea level to internal dynamics and surface forcing, excluding external gravity waves, atmospheric surface pressure response and tidal sea level changes.

The surface forcing is computed in an interactive way with 6-hourly ECMWF (European Center for Medium Range Weather Forecast) surface atmospheric analysis and forecast fields combined with the surface temperature of the model. The components of the surface net heat flux are computed on the basis of the bulk parameterizations of Reed (1977), for the surface solar radiation flux, of Bignami et al. (1995) for outgoing long-wave radiation and of Kondo (1975) for sensible and latent heat fluxes. The bulk formulation of Hellerman and Rosenstein (1983) is used in the wind stress computation. For more details of the implementation and tests of

the surface momentum and heat flux parameterizations see Castellari et al. (1998) and Angelucci et al. (1998). The surface heat flux derived from the atmospheric fields is corrected by relaxation of model SST to the observed satellite weekly SST. This method ensures that the model SST is close to the weekly mean SST. The corrected surface heat flux  $Q_C$ , used in the model simulations was computed as:

$$\frac{1}{\rho c_p} Q_C = \frac{1}{\rho c_p} Q + \lambda(T - SST), \quad (\text{B1})$$

where  $T$  is the model surface temperature,  $SST$  is the weekly observed field interpolated to the model time step between the previous week value and the present week and the nudging constant is equal to  $\lambda = 1.67$  m/day. The sea surface water flux is parameterized only with a salt flux given by the relaxation of model sea surface salinity toward a new climatology, called MED6 (Brankart and Pinardi, 2001). The relaxation constant is everywhere 2 m/day.

The model is initialized with a 7-year model experiment, forced with perpetual monthly mean forcing (a repeating seasonal cycle). The model is then run from 1 January 1997 to 1 September 1999 with surface forcing computed from ECMWF 6-h analyses. From 1 September 1999 to 4 January 2000 the model is run only with assimilation of XBT and SST heat flux correction. Then the MFSPP operational forecast period started 4 January 2000. From 4 April 2000 the combined assimilation of SLA and XBT was inserted into the operational forecasting procedure.

## Appendix C SOFA algorithm

### C.1 Optimal interpolation

Let us define the state vector  $\mathbf{x}_n$  as a vector formed by model state variables at the  $n$ -th model time step. Then the vector

$$\mathbf{x}_{n+1}^f = \mathbf{M}(\mathbf{x}_n^a) \quad (\text{C1})$$

is the model forecast, where  $\mathbf{M}$  is the model,  $\mathbf{x}_n^a$  is the analyzed estimate of the state vector at  $n$ -th time step. The vector of the observations  $\mathbf{y}_0$  is related to the true state  $\mathbf{x}^t$  by the equality:

$$\mathbf{y}_0 = \mathbf{H}(\mathbf{x}^t) + \epsilon \quad (\text{C2})$$

where  $\epsilon$  is the observational error and  $\mathbf{H}$  is the observation operator. Normally  $\mathbf{H}$  is a linear interpolation of model state variables into observational positions. In addition to that,  $\mathbf{H}$  may be more complex as explained in Appendix C.2 for satellite SLA. The forecast error covariance matrix is defined by:

$$\mathbf{B}_n^f = \mathbf{E} \left[ (\mathbf{x}_n^f - \mathbf{x}_n^t)(\mathbf{x}_n^f - \mathbf{x}_n^t)^T \right] \quad (\text{C3})$$

where  $E$  denotes the expectation operator. The analysis at each time step is computed as follows

$$\mathbf{x}_n^a = \mathbf{x}_n^f + \mathbf{K}(\mathbf{y}_0 - \mathbf{H}(\mathbf{x}_n^f)) \quad (\text{C4})$$

where  $\mathbf{K}$  is the Kalman gain. The Optimal interpolation is a particular case of the Extended Kalman Filter, when the forecast error covariance is replaced by background error covariance, Daley (1991):

$$\mathbf{B}_n^f = (\mathbf{D}_n^f)^{1/2} \mathbf{C} (\mathbf{D}_n^f)^{1/2} \quad (\text{C5})$$

where the background error variance diagonal matrix  $D^f$  is predicted and correlations  $\mathbf{C}$  are assumed to be constant. In our case  $\mathbf{C}$  is multivariate and contains cross correlations between  $T$ ,  $S$  and  $\psi$ , and also correlations between model state variables in horizontal and vertical directions. In the MFSPP data assimilation scheme  $\mathbf{D}_f$  was assumed to vary seasonally as explained later. The OI gain is defined as:

$$\mathbf{K}^{\text{OI}} = \mathbf{B}_n^f \mathbf{H}^T (\mathbf{H} \mathbf{B}_n^f \mathbf{H}^T + \mathbf{R})^{-1} \quad (\text{C6})$$

where  $\mathbf{R}$  is the data error covariance matrix.

SOFA is a reduced order multivariate optimal interpolation scheme. The order reduction is achieved by projecting the state vector into vertical EOFs which compose the columns of the  $\mathbf{S}$  matrix, the simplification operator. The vertical EOFs are the eigenfunctions of the  $\mathbf{B}_n^f$  matrix which is now written:

$$\mathbf{B}_n^f = \mathbf{S}^T \mathbf{B}_r^f \mathbf{S} \quad (\text{C7})$$

where  $\mathbf{B}_r^f$  contains the horizontal correlations of the  $n$  multivariate EOF modes defined in  $\mathbf{S}$ . Inserting Eq. (C7) into Eq. (C6) we obtain:

$$\mathbf{K} = \mathbf{S}^{-1} \mathbf{B}_r^f \mathbf{H}_r^T (\mathbf{H}_r \mathbf{B}_r^f \mathbf{H}_r^T + \mathbf{R}_r)^{-1} \quad (\text{C8})$$

where  $\mathbf{H}_r = \mathbf{H} \mathbf{S}^T$  and the matrix  $\mathbf{R}_r$  takes into account the representativity error in the reduced space.

Now Eq. (C6) can be rewritten as:

$$\mathbf{K}^{\text{OI}} = \mathbf{S}^{-1} \mathbf{K}^{\text{ROOI}} \quad (\text{C9})$$

where

$$\mathbf{K}^{\text{ROOI}} = \mathbf{B}_r^f \mathbf{H}_r^T (\mathbf{H}_r \mathbf{B}_r^f \mathbf{H}_r^T + \mathbf{R}_r)^{-1} \quad (\text{C10})$$

The order reduction procedure considers a limited number of vertical EOFs that should, however, still be representative of the error covariance matrix.

### C.2 Observational operator for the sea level anomaly

The observation operator for the sea level anomaly is calculated on the basis of the work of Pinardi et al. (1995). In addition to the linear interpolation, in this case it represents the combination of model state variables necessary to compute the observed SLA. In our case,  $\mathbf{H}$  contains only the geostrophic contribution to SLA, i.e.:

$$\eta = \frac{f\psi}{gH_0} - \frac{1}{\rho_0 H_0} \int_{-H_0}^0 (z' + H_0) \rho dz, \quad (\text{C11})$$



where the assumption is that the depth  $H_0$  is constant,  $\psi$  is the barotropic stream function,  $\rho_0$  is the reference density,  $\rho$  is the density,  $g$  – the gravity and  $f$  – the Coriolis parameter. In our case SLA assimilation is applied for the regions deeper than 1000 m. Correspondingly the parameter  $H_0$  is set equal to 1000 m.

## References

- Angelucci, M. G., Pinardi, N., and Castellari, S.: Air-sea fluxes from operational analysis fields: intercomparison between ECMWF and NCEP analysis over the Mediterranean area. *Phys. Chem. Earth*, 23, 569–574, 1998.
- Bignami, F., Marullo, S., Santoleri, R., and Schiano, M. E.: Long wave radiation budget in the Mediterranean Sea, *J. Geophys. Res.*, 100, 2501–2514, 1995.
- Brankart, J.-M. and Pinardi, N.: Abrupt cooling of the Mediterranean Levantine Intermediate water at the beginning of the eighties: observational evidence and model simulation, *J. Phys. Oceanogr.*, 31, 8, 2, 2307–2320, 2001.
- Buongiorno-Nardelli, B., Larnicol, G., D’Acunzo, E., Santoleri, R., 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 coupled patterns, *Ann. Geophysicae*, this issue, 2003.
- Castellari, S., Pinardi, N., and Leaman, K. D.: A model study of air-sea interactions in the Mediterranean Sea, *J. Mar. Syst.* 18, 89–114, 1998.
- Castellari, S., Pinardi, N., and Leaman, K.: Simulation of water mass formation processes in the Mediterranean Sea: Influence of the time frequency of the atmospheric forcing, *J. Geophys. Res.*, 105, 10, 24 157–24 181, 2000.
- Cox, M.: A primitive equation, 3-dimensional model of the ocean. GFDL Ocean Group Tech. Rep. 1, *Geophys. Fluid Dyn. Lab.*, Princeton, N. J., pp. 43, 1984.
- Daley, R.: *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, U.K., pp. 471, 1991.
- De Mey, P. and Benkiran, M.: A multivariate reduced-order optimal interpolation method and its application to the Mediterranean basin-scale circulation, In: *Ocean forecasting*, (Eds) Pinardi, N. and Woods, J., Springer Verlag, 281–306, 2002.
- De Mey, P. and Robinson, A.: Assimilation of altimetry data eddy fields in a limited-area quasi-geostrophic model, *J. Phys. Oceanogr.*, 17, 2279–2293, 1987.
- Demirov, E. and Pinardi, N.: Simulation of the Mediterranean Sea circulation from 1979 to 1993. Part I: The interannual variability, *J. Mar. Syst.*, 33/34, 23–50, 2002.
- Faucher, P., Gavart, M., and De Mey, P.: Isopycnal EOFs in the North and Tropical Atlantic and their use in estimation problems, *J. Geophys. Res.*, submitted, 2002.
- Hellerman, S. and Rosenstein, M.: Normal monthly wind stress over the world ocean with error estimates, *J. Phys. Oceanogr.*, 23, 1009–1039, 1983.
- Kondo, J.: Air-sea bulk transfer coefficients in diabatic conditions, *Boundary Layer Meteorol.*, 9, 91–112, 1975.
- Korres, G., Pinardi, N., and Lascaratos, A.: The ocean response to low frequency interannual atmospheric variability in the Mediterranean Sea. Part I: Sensitivity experiments and energy analysis, *J. Climate*, 13, 705–731, 2000.
- Le Traon, P. Y., Nadal, F., and Ducet, N.: An improved mapping Method of multisatellite altimeter data, *J. Atmos. Oceanic Tech.*, 15, 522–533, 1998.
- Le Traon, P. Y. and Ogor, F.: ERS-1/2 orbit improvement using Topex/Poseidon: the 2 cm challenge, *J. Geophys. Res.*, 103, 8045–8050, 1998.
- Manzella, G. M. R., Gardin, V., Cruzado, A., Fusco, G., Gacic, M., Galli, C., Gasparini, G. P., Gervais, T., Kovacevic, V., Millot, C., Petit DeLaVilleon, L., Spaggiari, G., Tonani, M., Tziavos, C., and Velasquez, Z.: EU-sponsored effort improves monitoring of circulation variability in the Mediterranean, *EOS Trans.*, AGU, 82, 43, 497–504, 2001.
- Murphy, A. H.: Skill score based on the mean square error and their relation to the correlation coefficient, *Monthly Weather Review*, 116, 2417–2424, 1988.
- Murphy, A. H.: What is a good forecast? An essay on the nature of goodness in weather forecasting, *Weather Forecasting*, 8, 281–293, 1993.
- Özsoy, E., Hecht, A., Ünlüata, Ü., Brenner, S., Sur, S., Bishop, H. I., Latif, M. A., Rozentroub, Z., and Oğuz, T.: A synthesis of the Levantine Basin circulation and hydrography, 1985–1990, in: *Physical oceanography of the Eastern Mediterranean Sea*, (Eds) Robinson, A. R. and Malanotte-Rizzoli P., *Deep-Sea Res.*, Part II, 40, 1075–1119, 1993.
- Pinardi, N., Rosati, A., and Pacanowski, R. C.: The sea surface pressure formulation of rigid lid models. Implications for altimetric data assimilation, *J. Marine Systems*, 6, 109–119, 1995.
- Pinardi, N. and Flemming, N.: *The Mediterranean Forecasting System Science Plan*, EuroGOOS Publication No. 11, Southampton Oceanographic Center, Southampton, 1998.
- Pinardi, N., Auclair, F., Cesarini, C., Demirov, E., Fonda-Umani, S., Gianni, M., Montanari, G., Oddo, P., Tonani, M., and Zavatareli, M.: Toward marine environmental predictions in the Mediterranean Sea coastal areas: a monitoring approach, in: *Ocean forecasting*, (Eds) Pinardi, N. and Woods, J., Springer Verlag., 339–376, 2002.
- Pinardi, N., Allen, I., Demirov, E., De Mey, P., Korres, G., Lascaratos, A., Le Traon, P. Y., Maillard, C., and Manzella G. M. R.: *The Mediterranean Ocean Forecasting System: First phase of implementation (1998–2000)*, *Ann. Geophysicae*, this issue, 2003.
- Raicich, F. and Rampazzo, A.: Observing System Simulation Experiments for the assessment of temperature sampling in the Mediterranean Sea, *Ann. Geophysicae*, this issue, 2003.
- Reed, R. K.: On estimation insolation over the ocean, *Progr. Oceanogr.*, 17, 854–871, 1977.
- Roussenov, V., Stanev, E., Artale, V., and Pinardi, N.: A seasonal model of the Mediterranean Sea circulation, *J. Geophys. Res.*, 100, 13 515–13 538, 1995.
- Sparnocchia, S., Pinardi, N., and Demirov, E.: Multivariate Empirical Orthogonal Function analysis of the upper thermocline structure of the Mediterranean Sea from observations and model simulations, *Ann. Geophysicae*, this issue, 2003.
- Walstad, L. J. and Robinson, A. R.: Hindcast and forecast of the POLYMODE data set with the Harvard Open-Ocean Model, *J. Phys. Oceanogr.*, 20, 1682–1702, 1990.