Ocean Science

Operational ocean models in the Adriatic Sea: a skill assessment

J. Chiggiato[1,*] **and P. Oddo**[2]

[1]Servizio IdroMeteorologico, ARPA Emilia Romagna, Viale Silvani 6, 40122 Bologna, Italy
[2]Istituto Nazionale di Geofisica e Vulcanologia, Unità Funzionale di Climatologia Dinamica, via Aldo Moro 44, 40127
Bologna, Italy
[*]now at: CNR-ISMAR, Castello 1364, 30122 Venezia, Italy

Abstract. In the framework of the Mediterranean Forecasting System (MFS) project, the performance of regional numerical ocean forecasting systems is assessed by means of model-model and model-data comparison. Three different operational systems considered in this study are: the Adriatic REGional Model (AREG); the Adriatic Regional Ocean Modelling System (AdriaROMS) and the Mediterranean Forecasting System General Circulation Model (MFS-GCM). AREG and AdriaROMS are regional implementations (with some dedicated variations) of POM and ROMS, respectively, while MFS-GCM is an OPA based system. The assessment is done through standard scores. In situ and remote sensing data are used to evaluate the system performance. In particular, a set of CTD measurements collected in the whole western Adriatic during January 2006 and one year of satellite derived sea surface temperature measurements (SST) allow to asses a full three-dimensional picture of the operational forecasting systems quality during January 2006 and to draw some preliminary considerations on the temporal fluctuation of scores estimated on surface quantities between summer 2005 and summer 2006.

The regional systems share a negative bias in simulated temperature and salinity. Nonetheless, they outperform the MFS-GCM in the shallowest locations. Results on amplitude and phase errors are improved in areas shallower than 50 m, while degraded in deeper locations, where major models deficiencies are related to vertical mixing overestimation. In a basin-wide overview, the two regional models show differences in the local displacement of errors. In addition, in locations where the regional models are mutually correlated, the aggregated mean squared error was found to be smaller, that is a useful outcome of having several operational systems in the same region.

Correspondence to: P. Oddo
(oddo@bo.ingv.it)

1 Introduction

Ocean physical processes play a crucial role in governing marine dynamics (acoustical, biological and sedimentological). Therefore, operational forecasting of physical ocean fields can greatly contribute to our understanding of the functioning of marine sub-systems, as well as providing an efficient support tool for marine environmental management (Oddo et al., 2006; Robinson and Sellschopp, 2002). For several applications such as fisheries management, naval operations, shipping, tourism, administration of marine resources and also for pure scientific purposes, high-resolution ocean forecasts are frequently required for limited regions (Onken et al., 2005).

Focusing on characteristic scales, processes and dynamics of a limited area, allows devoting particular attention to regionally specific numerical requirements (i.e. approximations, parameterizations, resolution and numerical techniques). Currently, several numerical models exist based on the same physical assumptions, and each single model shows its specific behaviour. Since model results derive from physical laws warped by numerical discretisation techniques, the possibility of having several numerical models implemented in the same area increases the confidence in model results.

In the framework of Mediterranean Forecasting System project (MFS, Pinardi et al., 2003) a suite of numerical ocean models has been developed and implemented in the Mediterranean Sea. The MFS modelling system is composed of a large-scale, coarse-resolution general circulation model covering the entire Mediterranean Sea (MFS-GCM) and a number of embedded high-resolution models in different regional seas. In the operational chain the MFS-GCM produces analysis/forecast for basin scale and provides initial and/or lateral boundary conditions for the regional models.

In this study the performance of the MFS modelling system in the Adriatic Sea is investigated. To our knowledge, two regional Operational Ocean Forecasting Systems

Fig. 1. Adriatic Sea coastline and bathymetry. Small circles represent CTDs locations. Note that iso-contours of depth 20 m, 50 m and 200 m, shown in the plot, are the very same used in the grouping done in Sect. 3. The grey thick line at 39 N represents AREG open boundary, while the black thick line at roughly 40 N represents AdriaROMS open boundary.

(hereinafter OOFS) are currently producing daily or weekly forecasts, published on free-access web sites, covering the whole Adriatic Sea with full three-dimensional implementation of the core ocean models: the Adriatic REGional forecasting system (AREG) and the Adriatic ROMS (AdriaROMS). The major aim of this work is to assess the performance of these two different regional OOFSs, eventually showing potential advantages deriving from specific regional implementation and from having more OOFSs in the same area. For completeness, the MFS-GCM is also included in this analysis, even if to a lesser extent, as a proxy of the relative large vs. regional scale systems performance. The analysis is focused on the quality of the operational systems, i.e., the agreement between model results and independent observations, therefore using the best available model output (best real-time estimates, including analyses). The relative skill to provide accurate short term forecast is left to other investigations.

The analysis is limited to temperature and salinity fields due to data availability.

The operational ocean forecasting systems are presented in Sect. 2. The comparisons between model results and in situ observations or remote sensing (AVHRR) data are presented and discussed in Sects. 3 and 4, respectively. Concluding remarks are given in Sect. 5.

2 Operational ocean forecasting systems

In this section all the considered operational forecasting systems for the comparison are briefly described. These systems differ in many aspects, such as their operational suite, spatial discretisation, physical parameterizations, and numerical weather prediction model used for the surface boundary condition. The major differences between operational forecasting systems are summarized in Table 1 for enhanced readability.

2.1 AREG

The AREG model domain covers the entire Adriatic Sea basin and extends into the Ionian Sea (Fig. 1). The horizontal resolution is approximately 5 km, while, for the vertical, 21 σ-coordinate levels are used. The model is based on the Princeton Ocean Model, POM (Blumberg and Mellor, 1987) as implemented in the Adriatic Sea by Zavatarelli and Pinardi (2003). The current implementation makes use of an iterative advection scheme for tracers (Smolarkiewicz, 1984) implemented into POM following Sannino et al. (2002). Surface forcing functions are obtained via standard bulk formulae (Oddo et al., 2005) using European Centre for Medium-range Weather Forecast (ECMWF) operational fields and AREG simulated sea surface temperature. A detailed description of the numerical model and forecasting system implementation can be found in Oddo et al. (2005, 2006).

2.2 AdriaROMS

AdriaROMS is the operational ocean forecast system for the Adriatic Sea running at the Hydro-meteorological Service of Emilia Romagna (ARPA-SIM). It is based on the Regional Ocean Modelling System (ROMS, detailed kernel description is in Shchepetkin and McWilliams, 2005). This Adriatic configuration has a variable horizontal resolution, ranging from 3 km in the north Adriatic to ∼10 km in the south, with 20 s-coordinate levels for the vertical. A third order upstream scheme is used for advection (Shchepetkin and McWilliams, 1998); a Laplacian operator adds a weak grid-size dependent on horizontal diffusivity, while no horizontal viscosity is used. The Mellor and Yamada (1982) 2.5 scheme is used for the vertical mixing, and density Jacobian scheme with spline reconstruction of the vertical profiles is used for the pressure gradient (Shchepetkin and McWilliams, 2003). The model was initialized in September 2004 from MFS-GCM fields optimally interpolated onto AdriaROMS grid, then run in pre-operational configuration until June 2005 when the first forecasts were published on the web.

Surface forcing is provided by the Limited Area Model Italy (LAMI, local implementation of the LM model, Steppeler et al., 2003), a non hydrostatic numerical weather prediction model with 7 km horizontal resolution providing trihourly shortwave radiation, 10 m wind, 2 m temperature,

Table 1. Summary of some of the most relevant differences between the three operational forecasting systems.

OOFS	MFS-GCM	AREG	AdriaROMS
Dataset	Analysis (weekly)	Hindcast (weekly)	Sequential forecast (03:00–24:00)
Horizontal resolution	1/16° (∼7 km)	5 km	Variable (3 km ÷ ∼10 km)
Vertical Resolution	72 uneven z-coordinate	21 sigma coordinate	20 non linear s-coordinate
Output	Daily averages	Daily averages	3-hourly snapshots
Initialisation	Summer 2004	Spring 2003	Fall 2004
Domain	Mediterranean Sea	Adriatic Sea	Adriatic Sea
Meteorological forcing	ECMWF analyses (1/2°, 6-hourly)	ECMWF analyses (1/2°, 6-hourly)	LAMI forecasts (7 km, 3-hourly)
Heat flux	Computed w/ flux correction (SST from AVHRR)	Computed w/out flux correction	Computed w/out flux correction
Fresh water flux	Relaxation to climatological SSS	Fresh water flux as salinity flux, all rivers but Po are climatological.	Only river flux (as source of mass and momentum); all rivers but Po are climatological
Data Assimilation	ARGO XBT SLA (only XBT in the Adriatic region)	none	none
Core Ocean Model	OPA	POM	ROMS

relative humidity, total cloud cover, mean sea level pressure and precipitation. All of them are used to compute momentum and heat fluxes. Long wave radiation is estimated using Berliand formula (Budyko, 1974), turbulent fluxes following Fairall et al. (1996), while no evaporation precipitation flux was included (added in a later version). MFS-GCM data are used at the open boundary to the south (see Fig. 1) with clamped boundary conditions with superimposed four major tidal harmonics (S2, M2, O1, K1), from the work of Cushman-Roisin and Naimie (2002), following Flather (1976). Forty-eight rivers and springs are included as well, using monthly climatological values from Raicich (1996). Persistence of daily discharge measured one day backward is used for the Po River.

2.3 MFS-GCM

The MFS-GCM (Tonani et al., 2008), based on the OPA code (Madec et al., 1998), covers the entire Mediterranean Sea with a horizontal resolution of 1/16 of degree and 72 un-evenly spaced z-coordinate levels on the vertical. The model is forced at the surface with ECMWF analysis and forecast atmospheric fields. It uses a reduced order optimal interpolation assimilation scheme (SOFA; De Mey and Benkiran, 2002; Demirov et al., 2003; Dobricic et al., 2007) to correct the model solution using vertical profiles from XBT and

ARGO, and satellite data of sea level anomaly (Pinardi et al., 2003), as well as flux corrections (relaxation to climatological sea surface salinity and sea surface temperature from AVHRR data). The ocean analysis-forecast consists of daily mean oceanographic fields computed for the entire Mediterranean basin. These fields are used in the two regional models to prescribe lateral open boundary conditions.

2.4 Selection of the time series

Databases of OOFSs output used in this paper are hindcast (AREG), forecast (AdriaROMS) and analysis (MFS-GCM). This choice followed from the OOFS operational chains. AREG and MFS-GCM produce weekly 9-day forecasts, thus combining a time series of weekly forecasts would also include the "day-of-the-week-dependent" signal of forecast growing error in the quality assessment. Selecting hindcast for AREG (analyses are not performed) and analyses for MFS-GCM allows building a time-series unaffected by this signal. On the other hand, forecasts obtained by AdriaROMS are produced daily (and no hindcasts or analyses are available) therefore a time series built using the first day of the forecast is the best dataset available. This approach is reasonable since it is the quality of the OOFS being assessed in this work, instead of the performance throughout the forecast range.
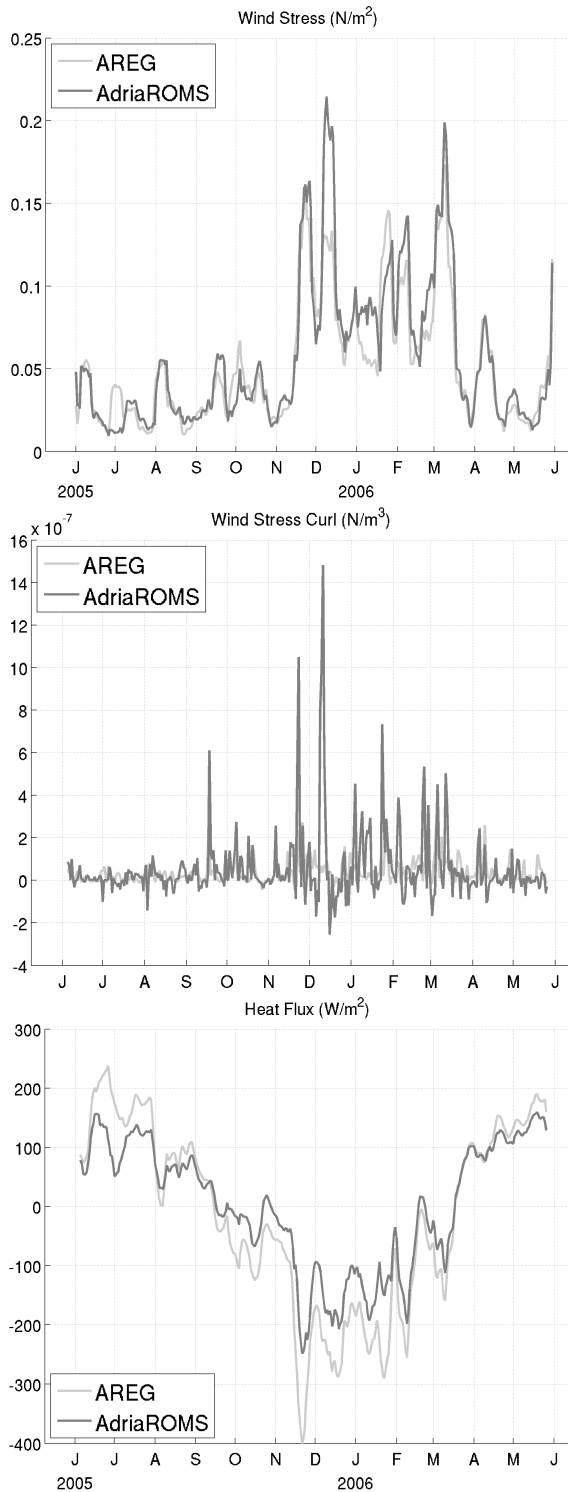
Fig. 2. Respectively from top to bottom: daily, basin averaged wind stress magnitude, wind stress curl and heat fluxes filtered with a 10-day running mean to enhance readability.

2.5 Diagnosed surface fluxes

Inter-comparison of the operational systems involves the relevant role of the atmospheric forcing; as given in Table 1, AREG and MFS-GCM are driven by a large-scale general circulation model (ECMWF), while AdriaROMS by a limited area model (LAMI). Due to the complex orography surrounding the Adriatic Sea, a proper representation of the coastal meteorology in that region is not trivial. Recent studies (Signell et al., 2005) have shown that in atmospheric models with coarse resolution (roughly larger than 10 km) wind fields are generally underestimated in the Adriatic, especially in the northern sector. Concerning heat fluxes, while the signal at the large scale may not necessarily be different using coarse or high resolution atmospheric models, some small-scale differences are expected due to the different physics and orography implemented in the models themselves.

The time series of wind stress magnitude and curl are shown in Fig. 2. Note that the basin-averaged magnitude is similar for both systems, while the wind stress curl shows similar patterns (mostly cyclonic vorticity) but often differs in magnitude, specially in winter. AdriaROMS has generally larger values due to steep gradients associated to limited area model small scale structures.

Major differences, even in the basin-averaged time series, are observed in the computation of surface heat fluxes. From a climatological point of view, the Adriatic Sea is believed to loose heat at the surface, although with large inter-annual variability and eventually years of heat gain (see Oddo et al., 2005 and citation therein). As seen in Fig. 2, AREG shows larger heat gain during summer, while AdriaROMS shows lesser heat loss during autumn-winter. On a short time scale, differences in the basin averaged between the two OOFSs can be as high as 100–150 W/m^2. In the considered year, AREG averaged heat flux is -14 W/m^2, associated to a meridional heat flux incoming from the southern boundary, while AdriaROMS heat flux is $+3$ W/m^2, with weak meridional heat flux outgoing through the southern boundary. Although AREG results seem more consistent with long term climatology (Artegiani et al., 1997), it is not possible to assess which behaviour is actually correct, given the limited observational dataset at hand.

Surface fresh water fluxes are not considered here since they are not routinely stored in the AdriaROMS nor AREG archive.

3 Comparison with in situ temperature and salinity

3.1 CTD data and methods

In January 2006, an extensive dataset of CTD measurements was collected during the cruise R/V URANIA in the western Adriatic Sea. This dataset (courtesy of CNR-ISAC Gruppo di Oceanografia da Satellite, Roma, IT) provided the

opportunity to assess a temporal snapshot of the ocean forecasting systems performance operating in the Adriatic Sea. The dataset consists of 150 CTD casts organized along 15 cross-shore sections (see Fig. 1), performed between 14 and 27 January 2006. The full CTD dataset has been split into four sub-categories, depending on the depth of the sampling positions. The grouping was done to evaluate scores from different coastal to open sea regions .The rationale is that regional forecasting systems are built to provide more accurate "information" on the coastal zones that may be roughly represented in large-scale systems; therefore, it is desirable to understand if the regional systems actually have skills in such areas.

The regions have been defined as follows:

1. Very shallow region (group G1): casts in depths not exceeding 20 m.

2. Shallow region (G2): casts in depths between 20 and 50 m.

3. Mid-depth region (G3): casts in depths between 50 and 200 m.

4. Deep region (G4): casts in depths exceeding 200 m.

A first general overview of the performance was done by means of Mean Errors (ME) and Root Mean Square Error (RMSE). Scores are estimated interpolating model result in time and space on the CTD locations. Within each group all squared errors have been aggregated over time and space (in a quasi-synoptic assumption) before taking the mean and the root mean square.

Rating the relative performance compared to MFS-GCM, it was decided to estimate the Mean Squared Error Skill Score (hereinafter MSESS, see Appendix A) at each CTD site and then to compute the number of significantly positive, significantly negative and not-significantly different values within each group. Therefore, a larger number of positive skill scores suggest a relatively better performance. Significance of the scores is estimated using the bootstrap technique with 1000 re-samples. Observations used in this assessment were not assimilated in the MFS-GCM.

3.2 Results

Results of the statistical assessment are summarized in Table 2. AdriaROMS shows negligible mean error in temperature in the two shallower groups, while temperatures are lower than those observed at G3 and, in particular, at G4, where there is a bias throughout the full water column (ME~RMSE). Given the depth at G4 (>200 m), and the time-scale of the major regional processes, this is unlikely due directly to surface forcing, but to the proximity of the open boundary and to the heat content derived by the initialization, since MFS-GCM itself is similarly biased in that region. AREG gives good accuracy for G1 and G4, while
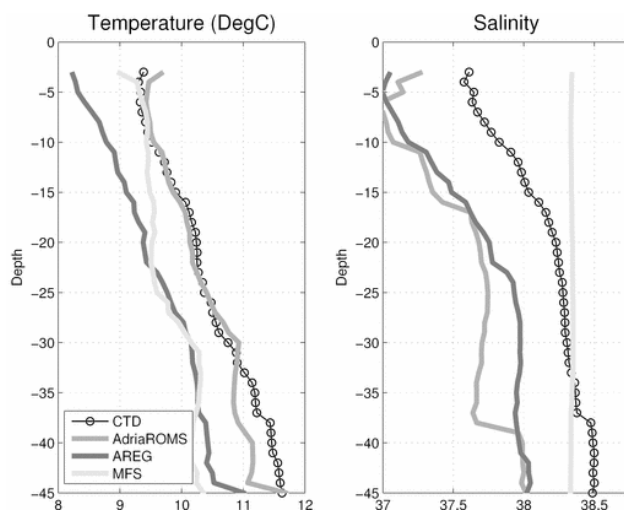
Fig. 3. Average profile for temperature (left panel) and salinity (right panel) of all the locations with CTD casts with maximum depth not exceeding 50 m.

lower temperatures in G2–G3. This behaviour is associated with horizontal diffusion problems, yielding the spreading of cold coastal waters inside the basin (Oddo et al., 2005). In general, even the temperature in MFS-GCM is lower than the observed temperature, with the exception of G1 due to the lack of the cold coastal water signal.

Analyzing the performance on salinity, both regional OOFSs have larger ME and RMSE in the very shallow group. This is easily explained by the difficulty to simulate the exact salinity in the western coastal current, since the models are anyway using climatological data for all rivers except the Po river. Errors decrease going toward deeper locations. Based on ME and RMSE, MFS-GCM is generally the most accurate.

Resulting MSESSs show that the regional systems have a larger number of positive scores (SK+ in Table 2) in group G1, regardless of the selected quantity. In the case of salinity, this may be counter-intuitive, since MFS-GCM has the lowest aggregated RMSE. Indeed, the RMSE is sensitive to extremes, and the aggregated RMSE for both AdriaROMS and AREG is downgraded by a low performance in a couple of locations only. Both regional systems generally show good performance in the representation of the very shallow coastal area as compared to larger-scale system. In the other groups the results are more regional model dependent, with AdriaROMS having greater accuracy in temperature, expect for G4, while AREG performs better in G4 and less in G2 and G3. The regional models do not perform better than the large-scale OOFS on salinity field (G1 excluded). The performance on MSESS of the regional models is mainly affected by biases. For example, the averaged vertical structures of AREG and AdriaROMS (in regions G1 and G2, see the panel on salinity Fig. 3) are consistent with the

Table 2. Mean error, root mean square error and mean square error skill scores in the four groups G1, G2, G3, G4, divided by the range of depth of the CTD locations. SK+, SK−, SK? mean, respectively, the number of profiles in which the skill score is significantly positive (regional system is best), negative (the large scale system is best) or not significantly different (neutral).
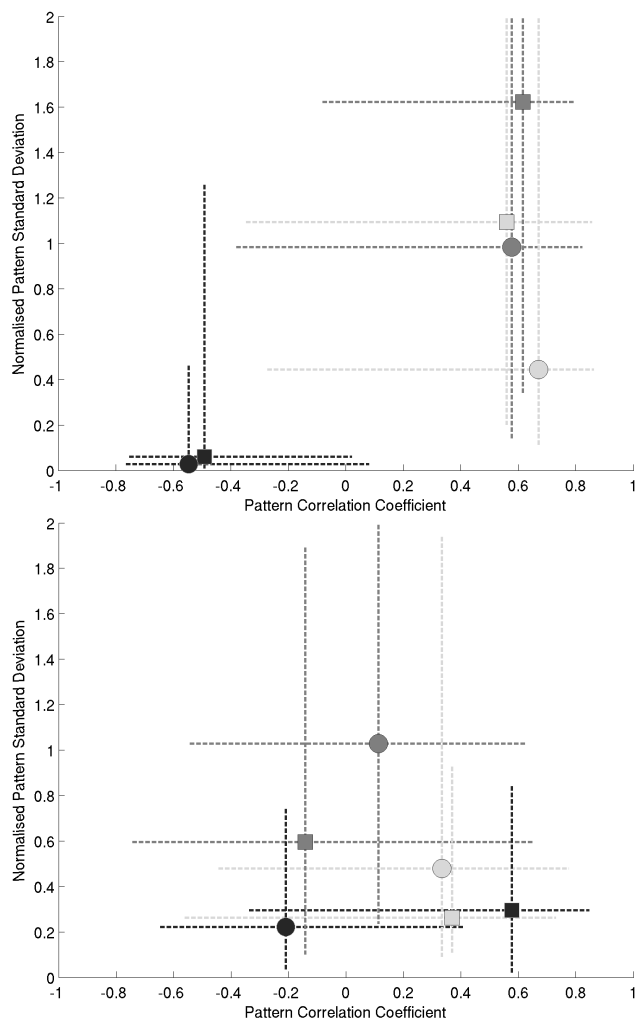
	TEMP					SALT				
	ME	RMSE	SK+	SK−	SK?	ME	RMSE	SK+	SK-	SK?
AdriaROMS										
G1 0–20	+0.09	1.03	16	4	1	−1.11	2.52	13	5	3
G2 20–50	−0.06	0.94	26	18	3	−0.54	0.72	4	40	3
G3 50–200	−0.73	0.98	30	26	2	−0.46	0.50	1	56	1
G4 200-inf	−1.05	1.09	1	22	0	−0.34	0.34	0	23	0
AREG										
G1 0–20	−0.03	1.35	11	7	3	−0.75	1.97	11	7	3
G2 20–50	−0.96	1.65	15	23	9	−0.48	0.85	12	32	3
G3 50–200	−0.90	1.32	18	37	3	−0.58	0.66	2	54	2
G4 200-inf	−0.02	0.25	16	7	0	−0.28	0.28	14	9	0
MFS										
G1 0–20	+1.13	1.97				+1.31	1.56			
G2 20–50	−0.73	1.49				+0.11	0.54			
G3 50–200	−0.76	1.25				−0.30	0.32			
G4 200-inf	−0.54	0.59				−0.31	0.31			

observations, giving good results on linear association and amplitude error, but biases eventually downgrade the performance on skill scores compared to MFS-GCM. It became then necessary to investigate further the actual skill of the regional OOFSs with other statistics, i.e., looking at amplitude and phase errors and not just at MSE. Results of these statistics are shown in Fig. 4a (grouping G1 and G2 together) and Fig. 4b (G3+G4). In G1+G2, the correlations in both AREG and AdriaROMS are similar and reasonably high for temperature and salinity (the medians are some 0.6, even if distributions are characterized by large spreading) with distribution of salinity in AREG and temperature in AdriaROMS centred on normalized standard deviation of unit (which is the most desirable value). Temperature in AREG instead often tends to overshoot the vertical stratification, while salinity in AdriaROMS to undershoot. The very low value of the medians for both normalized standard deviation and correlation in MFS-GCM suggest a low skill on reconstruction of the coastal gradient patterns, with the vertical profile being actually too homogeneous and often not even linearly and positively associated. On the other hand, looking at the sample distributions in deeper regions (G3+G4), the overall performance of the regional models is downgraded. The medians of the pattern correlation coefficients are now smaller in both AdriaROMS and AREG. MFS-GCM is more positively and linearly associated, at least, on temperature profiles (in this region XBT temperature data are assimilated in MFS-GCM). As a common feature, the OOFSs tend to underestimate the amplitude of profiles (with the exception of salin-

ity in AREG) predicting a too homogeneous vertical profile. Along with the sample distribution of the correlations, the two regional systems depict a lower skill in reproducing the vertical stratification in deeper regions, whereas the performance of the large scale system compares slightly better to its performance in the coastal locations, at least for the temperature. Deeper regions, affected by incoming Levantine waters (Artegiani et al., 1997) and, in general, by large scale circulation structures, make the regional OOFSs less effective in comparison to the large scale MFS-GCM system, which benefits also from the data assimilated.

3.3 Mutual and non-reciprocal behaviour between the operational systems
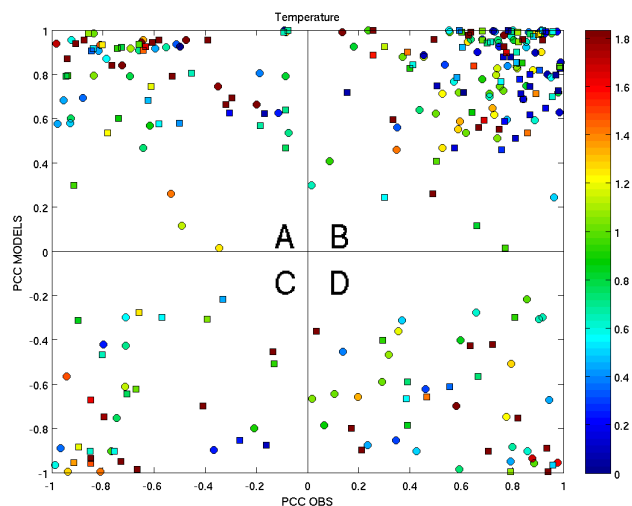
An additional analysis has been carried out to investigate the mutual behaviour of the two regional operational systems from a statistical point of view. For this aim, Pattern Correlation Coefficient between models (PCCm), between model and observations (PCCo) and RMSE have been combined, in this case grouping all the CTD casts. Results of the computation are shown in Fig. 5. On the basis of PCC values four areas (A, B, C and D) have been defined: area A (high PCCm and low PCCo values) identifies mutual systematic model errors; area B (high values of both PCCm and PCCo) indicates mutual skill; in area C (low values of both PCCm and PCCo) there are model-specific systematic errors and, finally, area D (low PCCm and high PCCo values) states for model-specific skill.

Fig. 5. The x- and y-axes indicate vertical integrated PCC with observation (PCCo) and between models (PCCm), respectively, while the colour indicates RMSE (models-observations) values. AdriaROMS values are indicated with circles, while AREG values are indicated with squares.
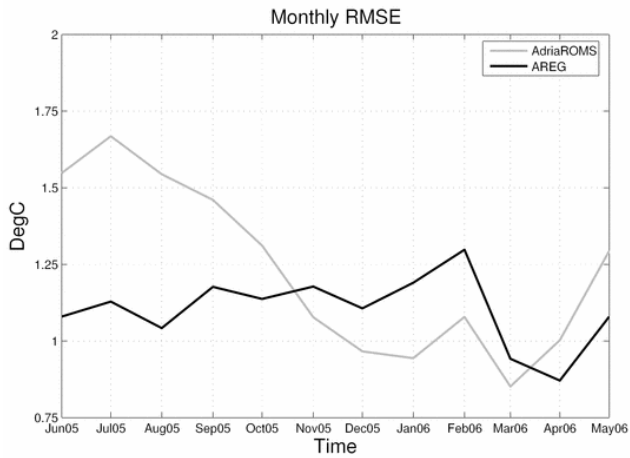
niques adopted. This suggest that efforts made to improve our knowledge, and consequently implementing the correct physics, will be more advantageous than improving the numerical techniques. The averaged RMSE in area B is 0.88, while in the region D it is about 1.16. Therefore, also considering only the model solution with positive PCCo (areas B-D) the portion of samples having positive PCCm is also generally a better estimation of the ocean state from a statistical point of view, i.e., having different models in the same region helps increasing the confidence of the results.
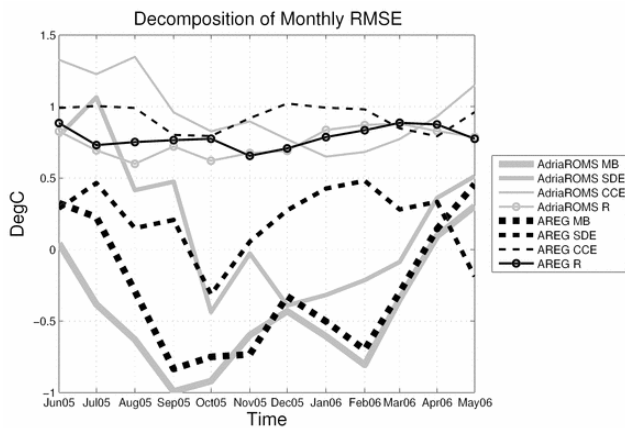
4 Comparison with AVHRR SST

In the previous section, the assessment of the selected operational products has been concentrated on a full three dimensional dataset, but covering a limited period of time. Operational sea surface temperature (SST) products, made available during MFSTEP and ADRICOSM projects lifetime, allow the assessment of a longer time-series, even if concerning a two-dimensional field only.

4.1 Data and method

The dataset consists of one year of AVHRR SST daily maps retrieved through the Pathfinder algorithm using composites of different night-time passages (Sciarra et al., 2006 and citation therein for details). Data are courtesy of CNR-ISAC, Gruppo di Oceanografia da Satellite, Rome (IT). SST data have been provided with clouds masked out and already mapped on AREG grid (approximately 5 km resolution) and for the sake of comparison, AdriaROMS results have been

Fig. 4. Sample distribution of normalized centred standard deviation and pattern correlation coefficient of model vs. observations in groups G1+G2 **(a)** and G3+G4 **(b)**. Squares depict the median of the distribution of temperature, circles the median in case of salinity. Dashed line shows the corresponding spread of 25th and 75th percentile. AdriaROMS is in light grey, AREG in medium grey, MFS-GCM in dark grey.

Considering both models, we found that 73% of total samples is within regions A and B (reciprocal performance), while 27% is in the regions C and D (non reciprocal performance). The averaged RMSE in A+B is 0.97 while in C+D is about 1.19 with a total average (A+B+C+D) of 1.03. Considering only model solutions having positive PCCm improve the systems quality in terms of RMSE.

In the systematic error macro-region (A+C) there is a total samples of 108 profiles with 72 in the A area and 36 within the C area. This means that most model errors derive from mutual models proprieties. For example, the error can depend on physical assumptions more than on numerical techniques.

Fig. 6. Time series of monthly averaged root mean square error of model vs. AVHRR-SST.


Fig. 7. Time series of monthly root mean squared error of model vs. AVHRR-SST decomposed in mean error term (MB), standard deviation error (STE), cross covariance error term (CCE) and correlation coefficient (*R*).

bilinearly interpolated onto the same grid. For equality in domains, this analysis is south-bounded at the latitude 40.7° N. MFS-GCM data are not considered in this analysis since the SSTs are used in the flux-correction procedure and therefore they are not an independent dataset.

The comparison between model SST and AVHRR skin SST may become critical when the warm layer develops and even worse in deeper regions (since model surface temperature is indeed representative of a thick layer, because of the terrain-following vertical coordinate). The fact that SST images are night-time images helps to minimize such biases.

Based on this dataset, the RMSE was computed monthly on a basin-wide, aggregated subset of model errors. Compared to other possible approaches, as for example first estimating over space the daily MSE and then averaging over time, this formulation permits overcoming the cloud-cover
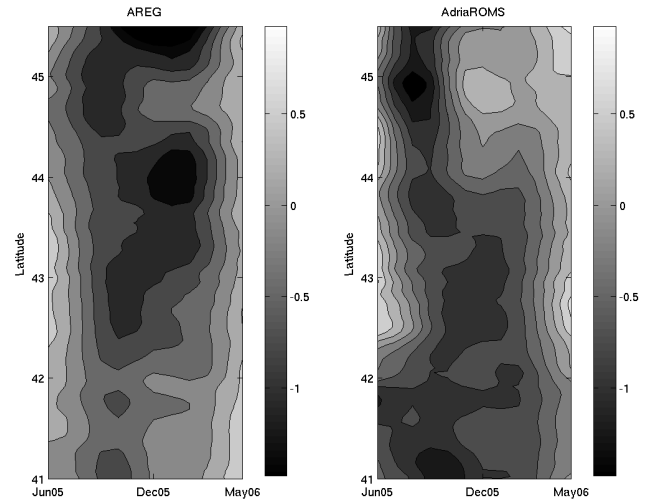
Fig. 8. Monthly zonal average of mean error between model (AREG, left panel and AdriaROMS, right panel) and AVHRR-SST. Units are °C.

problem (it gives a lesser weight to days with less spatial coverage).

4.2 Results

RMSEs estimated in the period June 2005–May 2006 are shown in Fig. 6. The two regional OOFSs have different scores, ranging approximately from 0.85 to 1.65°C. AdriaROMS has a large RMSE seasonal cycle with high summer values while better performance during winter. AREG shows similar values during summertime and wintertime, without a clear evidence of a seasonal cycle. In order to understand the source of the different intra-annual behaviours, the decomposition of the mean squared error was carried out following Oke et al. (2002) (see Appendix B for details).

Results of the decomposition are shown in Fig. 7. The basin-scale mean bias is roughly similar, except for summer 2005. Such difference can be partly explained by the use of daily averages as AREG SST, which would give a warm bias in summer. The correlation, ranging between 0.6–0.9, is roughly similar except for a few months. It seems, indeed, that the different seasonal behaviour is mostly controlled by the variations in the standard deviation, or amplitude error, which, in its turn, also affects the cross-correlation error (in the term $S_m S_o$ of Eq. (B1) in Appendix B) leading this term to a similar variability. The seasonal behaviour is then explained by the different amplitude of the sample distribution of the simulated temperature. In particular, clear causes for this different amplitude error are the cold water pool in the southern Adriatic in AdriaROMS during summer 2005 and the cold pools of water in AREG over the northern and middle Adriatic during winter. These pools can be seen in Fig. 8, where the zonally-averaged monthly mean bias is shown. In

general, the low performance in AdriaROMS in the southern region (which is also consistent with previous results in G4 for January 2006, see Sect. 3.2) is associated with a cold pool of water derived from the initialization and to an observed failure in the clamped-type open boundary conditions; in fact, when the dynamics of the nested model are not consistent with those of the outer model, spurious fronts develop preventing the correct flux through the boundary. This suggested switching from clamped boundary conditions to radiation-relaxation (Blayo and Debreu, 2005). In AREG, part of the errors (in particular those at latitude 44) are associated with horizontal diffusion problems (Oddo et al., 2005), yielding the spreading of cold coastal waters inside the basin. At the same time, the difference between the two systems in the northern region (latitude $>44°$) is also due to different heat fluxes. Actually, in the period from November 2005 to February 2006, AREG heat fluxes are, on the average, more negative by $44\,W/m^2$ (respectively, $-180\,W/m^2$ AREG and $-136\,W/m^2$ AdriaROMS with daily negative peaks of respectively, -459 and $-335\,W/m^2$) even if AREG SST is actually colder. These results for AREG are consistent with the performance found in G2–G3, as well as the good results in the southern Adriatic are consistent with the performance in G4 (see Sect. 3.2).

5 Summary and conclusions

Two regional operational forecasting systems, namely AREG and AdriaROMS, which were implemented during the MFS project lifetime and covering the whole Adriatic Sea region, have been presented and assessed. The forecasting systems differ in operational suite, core ocean model, parameterization of the physics as well as numeric, and meteorological forcing.

The performances of these operational systems have been evaluated by means of standard statistics. Available observations posed limitations to the assessment which is based on temperature and salinity only, not including any analysis on currents or other quantities.

With respect to the dataset of CTDs in January 2006, on the average, both systems provide simulated temperature and salinity lower than those observed. This result is mostly shared with the MFS-GCM, the large-scale system used here as benchmark of the performance.

The assessment based on mean squared error skill score outlines that the regional systems outperform the large-scale system in the very shallow region. Much of the performance in the offshore part of the basin is model- or region-dependent and eventually affected by bias. However, once biases are removed, the amplitude and linear association errors give better performance of the regional systems, that is, a higher skill in reconstructing vertical profiles. This is particularly true in the CTD locations not exceeding $50\,m$ depth, and only partly true in the deeper region, where, however, all

systems tend to have higher homogeneity along the vertical column compared to observations. Of course, the better performance in the coastal sector is a desirable outcome of the assessment, since regional systems are meant to add information in such challenging areas.

The systems show specific regions of low performance, and these areas are identified by both the assessment 3-D with CTDs and 2-D with satellite SST. Eventually these areas also regulate the different intra-annual cycle of RMSE in the SST field. However, based on the January 2006 data-set, these specific regions are only part of the story; in fact 67% of the CTDs with negative PPCo values are actually mutually shared by the models, and the remaining 33% represent only 11% of the total samples. At the same time, in locations where the regional models results are reciprocally correlated (73% of samples), the aggregated RMSE is lower with respect to the average value, suggesting an increase of confidence in model results. This is a positive outcome of having several operational systems in the same region, as well as the possibility of choosing the best model for a given area of interest. Besides, the availability of having continuous real-time results, with post processing techniques such as multi-model ensemble, may easily help in providing accurate real-time environmental pictures (Palmer et al., 2003; Krishnamurti et al., 2000).

Appendix A

The quality of OOFSs is usually assessed by means of basic statistics such as bias, root mean squared error (RMSE), pattern correlation coefficients and some skill scores (see Jolliffe and Stephenson, 2003, for a general review). Amongst these latter, one of the most used is probably the climatological skill score, or mean squared error skill score – MSESS (Murphy and Epstain, 1989), defined as follows:

$$\text{MSESS}=1-\frac{\frac{1}{n}\sum_{i=1}^{i=n}(m_i-o_i)^2}{\frac{1}{n}\sum_{i=1}^{i=n}(\text{ref}_i-o_i)^2} \tag{A1}$$

where m, o, ref, mean, respectively, model, observations and reference i-th value, n is the matched number of models-observations. In this case, the reference forecast can be in turn either climatology, persistence, or forecasts/analyses from another modelling system. For the sake of direct comparison between the regional and the large-scale systems, here the latter is used as reference in the MSESS estimates. Therefore, a positive (negative) skill score implies that the regional system is more (or less) skilful compared to the large-scale reference system.

Appendix B

Mean squared errors can be decomposed in many ways; following the Oke et al. (2002) approach and nomenclature, MSE can be split into:

$$MSE = MB^2 + SDE^2 + 2S_m S_o (1-CC) ; \qquad (B1)$$

where $MB = \bar{m} - \bar{o}$ is the mean bias, $SDE = S_m - S_o$ is the standard deviation error, $(2S_m S_o (1-CC))^{\frac{1}{2}}$ the cross correlation error, with m and o representing, respectively, model and observed values, S the standard deviation of the sample distribution, CC the correlation coefficient.

Edited by: E. J. M. Delhez

References

Artegiani, A., Bregant, D., Paschini, E., Pinardi, N., Raicich, F., and Russo, A.: The Adriatic Sea General Circulation, Part I: Air-Sea Interaction and Water Mass Structure, J. Phys. Oceanogr., 27, 1492–1514, 1997.

Blayo, E. and Debreu, L.: Revisiting open boundary conditions from the point of view of characteristic variables, Ocean Model., 9, 231–252, 2005.

Blumberg, A. F. and Mellor, G. L.: A description of a three-dimensional coastal ocean circulation model, in: Three-dimensional coastal ocean models, edited by: Heaps, N. S., American Geophysical Union, Washington D.C, 208 pp., 1987.

Budyko, K.: Climate and Life, Academic Press, 508 pp., 1974.

Cushman-Roisin, B. and Naimie, C. E.: A 3D finite-element model of the Adriatic tides, J. Mar. Syst., 37, 279–297, 2002.

DeMey, P. and Benkiran, M.: A multivariate Reduced-order Optimal Interpolation Method and its application to the Mediterranean Basin-scale Circulation, in: Ocean Forecasting, edited by: Pinardi, N. and Woods, J., 281–305, 2002.

Demirov, E., Pinardi, N., Fratianni, C., Tonani, M., Giacomelli, L., and DeMey, P.: Assimilation scheme of 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/.

Fairall, C. W., Bradley, E. F., Rogers, D. P., Edson, J. B., and Young, G. S.: Bulk parameterization of air-sea fluxes for Tropical Ocean Global Atmosphere Coupled-Ocean Atmosphere Response Experiment, J. Geophys. Res., 101(C2), 3747–3764, 1996.

Flather, R. A.: A tidal model of the northwest European continental shelf, Memories de la Societe Royale des Sciences de Liege, 6(10), 141–164, 1976.

Joliffe, I. T. and Stephenson, D. B.: Forecast Verification, A Practitioner's Guide in Atmospheric Science, John Wiley & Sons, 240 pp., 2003.

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bacjiochi, D., and Williford, E.: Multimodel Ensemble Forecast for Weather and Seasonal Climate, J. Climate, 13, 4196–4216, 2000.

Madec, G., Delecluse, P., Imbard, M., and Lévy, C.: OPA 8.1 Ocean General Circulation Model reference manual. Note du Pôle de modélisation, Institut Pierre-Simon Laplace, 11, 91 pp., 1998.

Mellor, G. L. and Yamada, T.: Development of a turbulence closure model for geophysical fluid problems, Rev. Geophys. Space Phys., 20, 851–875, 1982.

Murphy A. H. and Epstein, E. S.: Skill scores and correlation coefficients in model verification, Mon. Weather Rev., 119, 572–581, 1989.

Oddo, P., Pinardi, N., and Zavatarelli, M.: A numerical study of the interannual variability of the Adriatic Sea (2000–2002), Sci. Total Environ., 353, 39–56, 2005.

Oddo, P., Pinardi, N., Zavatarelli, M., and Coluccelli, A.: The Adriatic Basin Forecasting System, Acta Adriatica, 47 (Suppl.), 169–184, 2006.

Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D., Austin, J. A., Barth, J. A., Boyd, T. J., Kosro, P. M., and Levine, M. D.: A modelling study of the three-dimensional continental shelf circulation off Oregon. Part I: Model-Data Comparison, J. Phys. Oceanogr., 32, 1360–1382, 2002.

Onken, R., Robinson, A. R., Kantha, L., Lozano, C. J., Haley, P. J., and Carniel, S.: A rapid response nowcast/forecast system using multiply nested ocean models and distributed data systems, J. Mar. Syst., 56, 45–66, 2005.

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délécluse, P., Déqué, M., Díez, E., Doblas-Reyes, F. J., Feddersen, H., Graham, R., Gualdi, S., Guérémy, J. F., Hagedorn, R., Hoshen, M., Keenlyside, N., Latif, M., Lazar, A., Maisonnave, E., Marletto, V., Morse, A. P., Orfila, B., Rogel, P., Terres, J. M., and Thomson, M. C.: Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER), B. Am. Meteor. Soc., 85, 853–872, 2003.

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/.

Raicich, F.: On the fresh water balance of the Adriatic Sea, J. Mar. Syst., 9, 305–319, 1996.

Robinson, A. R. and Sellschopp, J.: Rapid assessment of the coastal ocean environment, in: Ocean Forecasting: Conceptual Basis and Applications, edited by: Pinardi, N. and Woods, J., Springer-

Verlag, NY, 199–229, 2002.

Sannino, G. M., Bargagli, A., and Artale, V.: Numerical modelling of the mean exchange through the Strait of Gibraltar, J. Geophys. Res., 107(C8), 3094, doi:10.1029/2001JC000929, 2002.

Sciarra, R., Bohm, E., D'Acunzo, E., and Santoleri, R.: The large scale observing system component of ADRICOSM: the satellite system, Acta Adriatica, 47 (Suppl.), 51–64, 2006.

Shchepetkin, A. and McWilliams J. C.: Quasi-monotone advection schemes based on explicit locally adaptive dissipation, Mon. Weather Rev., 126, 1541–1580, 1998.

Shchepetkin, A. F. and McWilliams, J. C.: A method for Computing Horizontal Pressure-Gradient Force in an Oceanic Model with a Non-Aligned Vertical Coordinate, J. Geophys. Res., 108(C3), 3090. doi:10.1029/2001JC001047, 2003.

Shchepetkin, A. F. and McWilliams, J. C.: The Regional Ocean Modelling System: A Split-Explicit, Free-Surface, Topography-Following-Coordinate Oceanic Model, Ocean Model., 9, 347–404, 2005.

Signell, R. P., Carniel, S., Cavalieri, L., Chiggiato, J., Doyle, J., Pullen, J., and Sclavo, M.: Assessment of wind quality for oceanographic modeling in semienclosed basins, J. Mar. Syst., 53, 217–233, 2005

Smolarkiewicz, P. K.: A fully multidimensional positive definite advection transport algorithm with small implicit diffusion, J. Comput. Phys., 54, 325–362, 1984.

Steppeler, J., Doms, G., Shatter, U., Bitzer, H. W., Gassmann, A., Damrath, U., and Gregoric, G.: Meso-gamma scale forecasts using the nonhydrostatic model LM, Meteorol. Atmos. Phys., 82, 75–96, 2003.

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/.

Zavatarelli, M. and Pinardi, N.: The Adriatic Sea Modelling System: a nested Approach, Ann. Geophys., 21, 345–364, 2003, http://www.ann-geophys.net/21/345/2003/.