Operational ocean models in the Adriatic Sea: a skill assessment

J. Chiggiato\textsuperscript{1} and P. Oddo\textsuperscript{2}

\textsuperscript{1}Servizio IdroMeteorologico, ARPA Emilia Romagna, Viale Silvani 6, 40122 Bologna, Italy

\textsuperscript{2}Istituto Nazionale di Geofisica e Vulcanologia, unità funzionale di Climatologia Dinamica, via Donato Creti 12, 40128 Bologna, Italy

Received: 30 November 2006 – Accepted: 8 December 2006 – Published: 21 December 2006

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

In the framework of the Mediterranean Forecasting System project (MFS) sub-regional and regional numerical ocean forecasting systems performance are assessed by mean of model-model and model-data comparison. Three different operational systems have been considered in this study: the Adriatic REGional Model (AREG); the AdriaROMS and the Mediterranean Forecasting System general circulation model (MFS model). AREG and AdriaROMS are regional implementations (with some dedicated variations) of POM (Blumberg and Mellor, 1987) and ROMS (Shchepetkin and McWilliams, 2005) respectively, while MFS model is based on OPA (Madec et al., 1998) code. The assessment has been done by means of standard scores. The data used for operational systems assessment derive from in-situ and remote sensing measurements. In particular a set of CTDs covering the whole western Adriatic, collected in January 2006, one year of SST from space born sensors and six months of buoy data. This allowed to have a full three-dimensional picture of the operational forecasting systems quality during January 2006 and some preliminary considerations on the temporal fluctuation of scores estimated on surface (or near surface) quantities between summer 2005 and summer 2006. In general, the regional models are found to be colder and fresher than observations. They eventually outperform the large scale model in the shallowest locations, as expected. Results on amplitude and phase errors are also much better in locations shallower than 50 m, while degraded in deeper locations, where the models tend to have a higher homogeneity along the vertical column compared to observations. In a basin-wide overview, the two regional models show some dissimilarities in the local displacement of errors, something suggested by the full three-dimensional picture depicted using CTDs, but also confirmed by the comparison with SSTs. In locations where the regional models are mutually correlated, the aggregated mean-square-error has been found to be lower, which is a useful outcome of having several operational systems in the same region.
1 Introduction

Ocean physical processes play a crucial role in governing marine dynamics (acoustical, biological and sedimentological). Therefore, operational forecasting physical ocean fields can greatly contribute to the 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, 2001). For several applications like fisheries management, naval operations, shipping, tourism, management of marine resources but also for pure scientific purposes fine resolution ocean forecasts are frequently required for limited regions (Onken et al., 2005).

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. A large scale, coarse resolution model covering the entire Mediterranean region (MFS model) and a number of embedded high-resolution models in different sub-regional seas compose the modelling system. The basic idea is to use the MFS model in order to produce analysis-forecast at basin scale and provide initial and/or lateral boundary conditions to the sub-regional models.

Obviously numerical models are not perfect and several decisions have to be taken by the scientists during the implementation phase (scale resolution, parameterisations and so on). Separating the scales of interest in the implementation phase (the decision to have specific model for different regions) allow to dedicate particular attention to regionally specific numerical requirements. Since the perfect model does not exist, also the perfect tuning is missing. At the present time, several numerical models exist based on the same physical assumptions, and each single model has its own behaviour. Since model results derive from physical laws warped by numerical discretisation techniques, the possibility to have several numerical models implemented in the same area increases the confidence in model results.

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

Due to data availability, the analysis has been limited to temperature and salinity
fields.

The operational ocean forecasting systems are presented in Sect. 2. In Sect. 3 the
comparison between model results and in situ observation is presented. In Sect. 4 the
comparison of model sea surface temperatures with remote sensing (AVHRR) data is
given. Finally, in Sect. 5, the conclusions of this work are summarised.

2 Operational ocean models

In this section all the operational forecasting systems considered in the comparison
are briefly described. These systems differ in many aspects, such as operational
suite, spatial discretisation, physical parameterisations, and numerical weather predic-
tion system used as surface boundary condition. For enhanced readability, the major
differences are also summarized in Table 1.
2.1 MFS

The MFS model (Tonani et al., 2006\textsuperscript{1}), based on the OPA code (Madec et al., 1998), covers the entire Mediterranean Sea with an horizontal resolution of 1/16° of degree and 72 unevenly spaced z-coordinate on the vertical. The model is forced at surface with European Centre for Medium range Weather Forecast (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., 2006) 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 in order to prescribe lateral open boundary conditions.

2.2 AREG

The AREG model domain covers the entire Adriatic Sea basin and extends into the Ionian Sea (Fig. 1). The horizontal resolution is approximately 5.0 km, while 21 terrain following $\sigma$-coordinate on the vertical. 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). A detailed description of the numerical model and forecasting system implementation can be found in Oddo et al. (2005, 2006).

2.3 AdriaROMS

AdriaROMS is the operational ocean forecast system for the Adriatic Sea running at ARPA-SIM. It is based on the Regional Ocean Modelling System (ROMS, detailed model 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 vertical terrain following coordinates. A third order upstream scheme is used for advection (Shchepetkin and McWilliams 1998); a laplacian operator adds a weak grid-size dependent horizontal diffusivity, while no horizontal viscosity is used. Mellor and Yamada (1982) scheme is used for the vertical mixing, and density jacobian with spline reconstruction of the vertical profiles is used for the pressure gradient (Shchepetkin and McWilliams 2003). The model has been initialised in September 2004 from MFS fields optimally interpolated onto AdriaROMS grid, then run in preoperational configuration until June 2005 when the first forecasts have been published on the web.

Surface forcing are provided by the atmospheric Limited Area Model Italy (LAMI, local implementation of the model LM, Steppeler et al., 2002), non hydrostatic NWP model with 7 km horizontal resolution, that provides tri-hourly shortwave radiation, 10 m wind, 2 m temperature, 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 M. E. and T. G.: Berliand formula (Budyko, 1974), turbulent fluxes following Fairall et al. (1996) while no evaporation precipitation flux was included (added in a later version). MFS data are used at the open boundary to the south (see Fig. 1) with clamped boundary conditions (by the way, switched to relaxation-radiation following Blayo and Debreu (2005) after the time period considered in this work) 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 value from Raicich (1996). For the Po River instead, persistence of the daily discharge measured one day backward is used.
3 Comparison with in situ temperature and salinity

In January 2006, an extensive dataset of CTD measurements has been collected during the cruise done with R/V URANIA, covering the western part of the Adriatic Sea. This dataset (courtesy of ISAC-CNR Gruppo di Oceanografia da Satellite, Roma, IT) has 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 organised along 15 cross-shore sections (see Fig. 1), performed between 14 and 27 January 2006.

The full CTD dataset has been split in four sub-categories, depending on the depth of the sampling positions. The grouping has been done in order to evaluate scores in different regions (from coastal to open sea). In general, regional forecasting systems are built to provide more accurate “information” in the coastal zones that may be crudely represented in the large-scale system, therefore it is desirable understand if the regional systems add skills in such areas.

The regions have been defined as follows:

1. Very shallow region (group G1): casts on depths not exceeding 20 m.
2. Shallow region (G2): casts on depths between 20 and 50 m.
3. Mid-depth region (G3): casts on depths between 50 and 200 m.
4. Deep region (G4); casts on depths exceeding 200 m.

The quality of OOFS is usually assessed by means of basic statistics such as bias, root mean squared error (RMSE), 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 Epstein, 1989), defined as follows:

\[
\text{MSESS} = 1 - \frac{1}{n} \sum_{i=1}^{i=n} \frac{(m_i - o_i)^2}{(ref_i - o_i)^2}
\]

(1)

where \(m\), \(o\), \(ref\), mean respectively model, observations and reference ith value, \(n\) the matched number of models-observations. In this case, the reference forecast in turn can be a climatology, persistence, or forecasts/analyses from another modelling system. For 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 (less) skilful compared to the large-scale reference system.

The comparison may appear somewhat unfair for the regional systems. MFS model in fact does have data assimilation and flux correction (Demirov and Pinardi, 2002; Dobricic et al., 2006; Tonani et al., 2006) which should prevent drifts. The regional systems instead (continuous forecast and continuous hindcast) are free to evolve (and drift). On the other hand, a better performance of the latter at least on the coastal zone is still desirable.

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

Rating the relative performance compared to MFS, it has been chosen to estimate the skill score in each CTD location 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 score suggest a relatively better performance. Significance of the score has been estimated using bootstrapping technique with 1000 re-samples. Results are summarised in Table 2.

Basically, with a few exceptions, all the models are fresher and colder than observations.

AdriaROMS shows a temperature mean RMSE of some 1°C, irrespective the group. The ME instead has larger magnitude going toward deeper regions. AREG shows a mean RMSE higher than 1.3°C in G1-G2-G3, while sensibly lower in G4. Better MEs are found in the very shallow and in the deep region. MFS has a similar behaviour, but largest errors on the group G1.

Analysing the performance on salinity, the models 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. Reasonably, errors get lower going toward deeper locations.

The comparison by means of MSESS shows that the regional systems have a larger number of positive scores in group G1, irrespective to the selected quantity. In the case of salinity, this may be counterintuitive since MFS has the lowest aggregated RMSE. Indeed, the RMSE is sensitive to extrema, and the aggregated RMSE for both AdriaROMS and AREG is somewhat downgraded by a low performance on a couple of locations only. In general, both the regional systems show a good performance in the representation of the challenging very shallow coastal area, compared to the large scale system.

In the other groups the results are somewhat more regional model dependent.

AdriaROMS performs better in temperature only, and excluded G4. G4 is somewhat critical for this system, since this region (basically the deep southern Adriatic) is very close to the lateral open boundary and has a spatial resolution of about 10 km (coarser than MFS model).

AREG performs largely better in G4, while much less in G2-G3 (somewhat an oppo-
site behaviour between the two regional OOFSs).

Much of the performance depicted by the MSESSs in the coastal areas is likely associated to biases. An example is provided in Fig. 2. The averaged vertical structures of AREG and AdriaROMS are consistent with the observations, leading to good results on linear association and amplitude error. On the other hand, the bias may be higher eventually downgrading the performance on skill scores.

On the other hand, amplitude and phase errors of the regional OOFSs in the coastal CTD locations are different compared to those in deeper waters. A flavour of this behaviour is show in Fig. 3a (grouping G1 and G2 together) and Fig. 3b (G3+G4). The figures show the median, the 25th and 75th percentiles of the sample distribution of the centred (i.e., bias removed) standard deviation normalised by the standard deviation of the observations and the pattern correlation coefficients (PCC). Regarding 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 the distributions is characterised by large spreading) with distribution of salinity in AREG and temperature in AdriaROMS centred on normalised 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 normalised standard deviation and correlation in MSF suggest a low skill on reconstruction of the coastal gradients pattern, with the vertical profile being actually too homogeneous and often not even linearly positively associated. 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 lower in AdriaROMS and even more in AREG. Now MFS model is instead more positively linearly associated at least on temperature profiles (in this region XBT temperature data are assimilated). As a common feature, all the OOFSs tend to underestimate the amplitude of the profiles (with the exception of salinity in AREG), that is, 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
An additional example of the ability of the regional models to reproduce coastal gradients can be drawn by means of a comparison with buoy data available near the Po river mouth (Fig. 1), one of the most challenging locations from the point of view of variability in salinity. This buoy, namely S1, provides hourly temperature and salinity data below the surface (0.5 m depth).

The comparison has been done on daily averages, therefore averaging to daily both buoy data and 3-hourly AdriaROMS results (AREG and MFS are already daily) before computing any statistics. Results are presented in Fig. 4. Beyond the obvious better performance of the regional models (the large-scale models do not have any river implemented), it is noticeable the negligible ME even if with relatively high scatter (see Table 3 for ME and RMSE values). This behaviour is not found out for temperature, the three model performing nearly the same way (plot not shown, see Table 3 for statistics).

In order to investigate the advantage of having different regional OOFS in the same area and to synthesise the operational systems results an additional analysis has been carried out combining PCC between models (PCCm), between model and observations (PCCo) and RMSE, grouping all the CTD casts. Results of the computation are shown in Fig. 5. On the base of PCC values four areas (A, B, C and D) have been defined: area A (high PCCm and low of 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. Some macro-proprieties of this analysis are: the two models have identical distribution between areas A and B; B is the most populated region while C is the area with less data.
Considering both the models, we found that 73% of total population is within region A and B, while 27% is in the regions C and D. 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 1.17. We can state that considering only the model solutions having positive PCCm improve the systems quality results of 17% in term of RMSE.

In the systematic error macro-region (A+C) there is a total population of 108 with 72 in the A area and 36 within C area. This means that most of the models error derives from mutual models proprieties. For example, the error can derive from physical assumption more than numerical technique used. This suggests that spending effort in improving the knowledge, and as consequence implement the correct physic, will give more advantage than improve 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 this population having positive PCCm also is generally a better estimation of the ocean state.

The 67% of model results with positive PCCm are characterized also by positive PCCo values, on the contrary only the 55% of models results with negative PPCm values show positive correlation with the observations. Having different models in the same region increase the confidence of models results also in terms of PCC.

4 Comparison with AVHRR SST

The observational dataset used in this section consists of one year of AVHRR composites providing sea surface temperature (SST). A daily SST map has been retrieved through the Pathfinder algorithm using composites of different night-time passages (Sciarrà et al., 2006, and citation therein for details). Data are courtesy of ISAC-CNR Gruppo di Oceanografia da Satellite, Rome (IT). SST data have been provided with clouds masked out and already mapped on AREG grid (approximatively 5 km resolution) and for sake of comparison, AdriaROMS results have been bilinearly interpolated...
onto the same grid. For equality of domains, this analysis has been limited to the south to the latitude 40.7 N.

The comparison between model SSTs and AVHRR skin SST may be critical when the warm layer develops and even worse in deeper regions (where the 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 does anyway help to minimise such biases.

MFS data are not considered in this analysis since these SSTs are used in the flux-correction procedure and therefore is not an independent dataset.

Based on this dataset, the RMSE has been computed on monthly, basin-wide, aggregated subset of model outputs-observations. Compared to other possible approaches, for example first estimating over space the daily MSE and then averaging over time, this formulation permits to overcome the cloud-cover problem (it gives a lesser weight to days with less spatial coverage).

The RMSE estimated in the period June 2005–May 2006 are shown in Fig. 6. The two regional OOFSs present indeed different scores, ranging roughly between 0.85 and 1.65°C. AdriaROMS has a large RMSE seasonal cycle with high summer values while much better performance during winter. AREG shows better values during summertime compared to wintertime, without a clear evidence of a seasonal cycle. In order to understand the source of the difference of this intra-annual behaviour, the decomposition of the mean squared error has been carried out, following Oke et al. (2002) approach and nomenclature that is:

\[ MSE = MB^2 + SDE^2 + 2S_mS_o (1 - CC); \] (2)

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

Results of the decomposition are shown in Fig. 7. It can be seen that the basin-
scale mean bias is roughly similar, except summer 2005. Such difference can be, at least 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 in a few months. It seems indeed that the different seasonal behaviour is mostly controlled by the variations in the standard deviation error, or amplitude error, which in its turn acts also in the cross-correlation error (in the term \( S_m S_o \)) leading this term to a similar variability. Basically, the different seasonal behaviour is then associated to a larger range of large scale variability in AdriaROMS in summer, while larger in AREG in winter.

The fact that the mean bias between models is roughly similar is somewhat by chance: in an intra-basin variability it is indeed different in magnitude, while the annual cycle (maximum in spring and minimum in fall) still holds. This can be noted in Fig. 8, where the monthly, zonal averaged mean bias is presented. The mean bias latitudinal displacement between the two models is in fact opposite. In AREG the main source of most negative biases is in the mid-northern part, in AdriaROMS is instead the mid-southern, yet with a some positive biases in the northern. The reasons for that difference are potentially many, first of all the role of the different meteorological forcing, and only a dedicated process-study can elucidate the relative role of such driving mechanisms. In general, in AdriaROMS, the low performance in the southern region (which, by the way, is also consistent with previous results in G4 for January 2006) is somewhat associated to a failure in the open boundary conditions, which posed the necessity to switch from clamped boundary conditions to radiation-relaxation, following Blayo and Debreu (2005). In AREG, part of the errors (in particular those at the latitude 44) are associated to horizontal diffusion problems (Oddo et al., 2005), yielding the spreading of cold coastal waters inside the basin. These results are again somewhat consistent with the performance in G2-G3, as well as the good results in the southern Adriatic which is consistent with the performance in G4.
5 Summary and conclusions

The regional operational forecasting numerical models, namely the AREG and AdriaROMS systems, which cover the full Adriatic Sea region implemented during MFS project lifetime, have been presented and compared with the MFS model covering the entire Mediterranean basin and providing lateral boundary conditions data for the regional models.

All the forecasting systems differ in operational suite, core ocean model, parameterisation of the physics and meteorological forcing.

These operational systems performances have been evaluated by means of data-model and model-model comparison using standard statistics. Available observations posed the limits of the comparison, which is in fact based on temperature and salinity and does not include any analysis on currents or other quantities. On the other hand, the dataset of observations include many source of data, such as CTDs, remotely sensed sea surface temperature and data from one coastal buoy. This allowed having a full three-dimensional picture of the operational forecasting systems quality during January 2006 and some preliminary considerations on the temporal fluctuation of scores estimated on surface or near surface quantities between summer 2005 and summer 2006.

Within the bounds posed by available observations, in the three-dimensional snapshot of January 2006 the regional models has been found to be colder and fresher than observations. Reasonably, both eventually outperform the large scale model in the shallowest locations. Performance on amplitude and phase errors are also much better in locations shallower than 50 m, while degraded in deepest locations, where the models tends to have a higher homogeneity along the vertical column compared to observations.

In a basin-wide overview, the two regional models show somewhat opposite local displacement of performances; AdriaROMS outperform the other models in temperature excluding the deepest, southern Adriatic, region, where AREG shows instead better
results (as well as in salinity). This is also somewhat confirmed also by the comparison with AVHRR SST.

At the same time, in locations where the regional models are mutually correlated, the aggregated RMSE has been found to be lower, which is a positive outcome of having several operational systems in the same region, as well as positive outcome is the possibility to choose the best model given a certain area of interest, and nonetheless, the availability of having continuous real-time results which post processing procedure such as multi-model ensemble techniques may easily improve providing accurate real-time environmental pictures (Palmer et al., 2003; Krishnamurti et al., 2000). This work provided insights for the revision of the operational systems considered.

Acknowledgements. The authors greatly acknowledge ISAC-CNR Gruppo di Oceanografia da Satellite for AVHRR and CTD data, ISMAR CNR Bologna for buoy S1 data. This work has been carried out within the framework of EU VFP MFSTEP project, as well as ADRICOSM project funded by the Italian Ministry of Environment and Territory (through the University of Bologna, Centro Interdipartimentale per le Ricerche di Scienze Ambientali, Ravenna, Italy) and European INTERREG III CADSES – CADSELAND project. J. Chiggiato also acknowledge H. Arango, J. Wilkin (Rutgers University) and R. P. Signell, J. Warner (USGS) for useful hints during the implementation phase of AdriaROMS.

References

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


De Mey, P. and Benkiran, M.: A multivariate Reduced-order Optimal Interpolation Method and


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


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

<table>
<thead>
<tr>
<th>OOFS</th>
<th>MFS</th>
<th>AREG</th>
<th>AdriaROMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>Analysis (weekly)</td>
<td>Hindcast (weekly)</td>
<td>Sequential forecast (03:00–24:00)</td>
</tr>
<tr>
<td>Horizontal resolution</td>
<td>1/16° (~7 km)</td>
<td>5 km</td>
<td>Variable (3 km ÷~10 km)</td>
</tr>
<tr>
<td>Vertical Resolution</td>
<td>72 uneven z-coordinate</td>
<td>21 sigma coordinate</td>
<td>20 non linear s-coordinate</td>
</tr>
<tr>
<td>Output</td>
<td>Daily averages</td>
<td>Daily averages</td>
<td>3-hourly snapshots</td>
</tr>
<tr>
<td>Initialisation</td>
<td>Summer 2004</td>
<td>Spring 2003</td>
<td>Fall 2004</td>
</tr>
<tr>
<td>Domain</td>
<td>Mediterranean Sea</td>
<td>Adriatic Sea</td>
<td>Adriatic Sea</td>
</tr>
<tr>
<td>Meteorological forcing</td>
<td>ECMWF analyses (1/2°, 6-hourly)</td>
<td>ECMWF analyses (1/2°, 6-hourly)</td>
<td>LAMI forecasts (7 km, 3-hourly)</td>
</tr>
<tr>
<td>Heat flux</td>
<td>Computed w/flux correction (SST from AVHRR)</td>
<td>Computed w/out flux correction</td>
<td>Computed w/out flux correction</td>
</tr>
<tr>
<td>Fresh water flux</td>
<td>Relaxation to climatological SSS</td>
<td>Fresh water flux as salinity flux, all rivers but Po are climatological.</td>
<td>Only river flux (as source of mass and momentum); all rivers but Po are climatological none</td>
</tr>
<tr>
<td>Data Assimilation</td>
<td>ARGO XBT SLA (only XBT in the Adriatic region)</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Core Ocean Model</td>
<td>OPA</td>
<td>POM</td>
<td>ROMS</td>
</tr>
</tbody>
</table>
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 (better the regional system), negative (better the large scale system) or not significantly different (neutral).

<table>
<thead>
<tr>
<th></th>
<th>ME</th>
<th>RMSE</th>
<th>SK+</th>
<th>SK-</th>
<th>SK?</th>
<th>ME</th>
<th>RMSE</th>
<th>SK+</th>
<th>SK-</th>
<th>SK?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AdriaROMS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>0.09</td>
<td>1.03</td>
<td>16</td>
<td>4</td>
<td>1</td>
<td>-1.11</td>
<td>2.52</td>
<td>13</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>G2</td>
<td>-0.06</td>
<td>0.94</td>
<td>26</td>
<td>18</td>
<td>3</td>
<td>-0.54</td>
<td>0.72</td>
<td>4</td>
<td>40</td>
<td>3</td>
</tr>
<tr>
<td>G3</td>
<td>-0.73</td>
<td>0.98</td>
<td>30</td>
<td>26</td>
<td>2</td>
<td>-0.46</td>
<td>0.50</td>
<td>1</td>
<td>56</td>
<td>1</td>
</tr>
<tr>
<td>G4</td>
<td>-1.05</td>
<td>1.09</td>
<td>1</td>
<td>22</td>
<td>0</td>
<td>-0.34</td>
<td>0.34</td>
<td>0</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td><strong>AREG</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>-0.03</td>
<td>1.35</td>
<td>11</td>
<td>7</td>
<td>3</td>
<td>-0.75</td>
<td>1.97</td>
<td>11</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>G2</td>
<td>-0.96</td>
<td>1.65</td>
<td>15</td>
<td>23</td>
<td>9</td>
<td>-0.48</td>
<td>0.85</td>
<td>12</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td>G3</td>
<td>-0.90</td>
<td>1.32</td>
<td>18</td>
<td>37</td>
<td>3</td>
<td>-0.58</td>
<td>0.66</td>
<td>2</td>
<td>54</td>
<td>2</td>
</tr>
<tr>
<td>G4</td>
<td>-0.02</td>
<td>0.25</td>
<td>16</td>
<td>7</td>
<td>0</td>
<td>-0.28</td>
<td>0.28</td>
<td>14</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td><strong>MFS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>+1.13</td>
<td>1.97</td>
<td></td>
<td></td>
<td></td>
<td>+1.31</td>
<td>1.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>-0.73</td>
<td>1.49</td>
<td></td>
<td></td>
<td></td>
<td>+0.11</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>-0.76</td>
<td>1.25</td>
<td></td>
<td></td>
<td></td>
<td>-0.30</td>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>-0.54</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td>-0.31</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Root mean square error and mean error of daily averages of temperature and salinity at buoy S1 (respectively, TS1 and SS1). The period considered is from 1 July to 31 December 2005.

<table>
<thead>
<tr>
<th></th>
<th>AdriaROMS</th>
<th>AREG</th>
<th>MFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>TS1 0.01</td>
<td>−0.27</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>SS1 0.29</td>
<td>−0.17</td>
<td>4.82</td>
</tr>
<tr>
<td>RMSE</td>
<td>TS1 1.23</td>
<td>0.93</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>SS1 2.73</td>
<td>2.41</td>
<td>5.95</td>
</tr>
</tbody>
</table>
Fig. 1. Adriatic Sea coastline and bathymetry. Location of the measurements used are also shown: small circle are CTDs locations, light-grey square is S1 buoy. Note that iso-contour of depth 20 m, 50 m and 200 m, shown in the plot, are those used in the grouping done in Sect. 3.
Fig. 2. Average profile for temperature (left panel) and salinity (right panel) of all the locations with CTD casts with maximum depth not exceeding 50 m.
Fig. 3. Sample distribution of normalized centred standard deviation and pattern correlation coefficient of model vs. observations in group G1+G2 (a) and G3+G4 (b). Squares depict the median of the distribution of temperature, circles the median in case of salinity. Dashed lines shows the corresponding spread of 25th and 75th percentile. AdriaROMS is in light grey, AREG in medium grey, MFS in dark grey.
Fig. 4. Time series of daily averages salinity at S1 buoy location at 0.5 depth.
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 shown with circle makers while AREG values with squared markers.
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).
Fig. 8. Monthly zonal average of mean error between model (AREG, left panel and AdriaROMS, right panel) and AVHRR-SST. Units are °C.