Project of Strategic Interest NextData

Deliverable D1.3.1
Historical marine data and quality control procedure for the Mediterranean Sea Reconstruction/Reanalysis (RR)

Resp: Nadia Pinardi, INGV

Introduction

A feasibility study for a Reconstruction/Reanalysis (hereafter RR) of the Mediterranean Sea for the past hundred years has been carried out during the first year of the project. One of the main tasks was the **review, update and collection of the existing historical marine data for the RR period 1912-2011**. Observations are a fundamental component of the Mediterranean RR system, as schematized in Figure 1. An integrated observing and prediction system relies on the observational component in different phases:

- model initialization
- model calibration
- data assimilation
- validation of model results and quality assessment of the RR products.

The RR system needs in situ temperature and salinity observations along the water column and remote sensed observations of altimetry (SLA) to be assimilated in the numerical model (Figure 1). Satellite Sea Surface Temperature (SST) observations are not assimilated in the RR system but they are used to correct the surface heat flux on the base of the difference between modeled SST and observed SST.

In situ temperature and salinity profiles are also used to compute statistical gridded products, like monthly climatology, to initialize the RR system but also for future quality assessment of RR results.

A brief description of remote sensed observations is given and further details are included in Deliverable D2.2.1. No Quality Control (QC) procedure has been implemented for this kind of data since their quality assessment has been already carried out and documented from the data provider. It will follow a description of the in situ data set (see also D.2.2.1) and the QC strategy adopted.

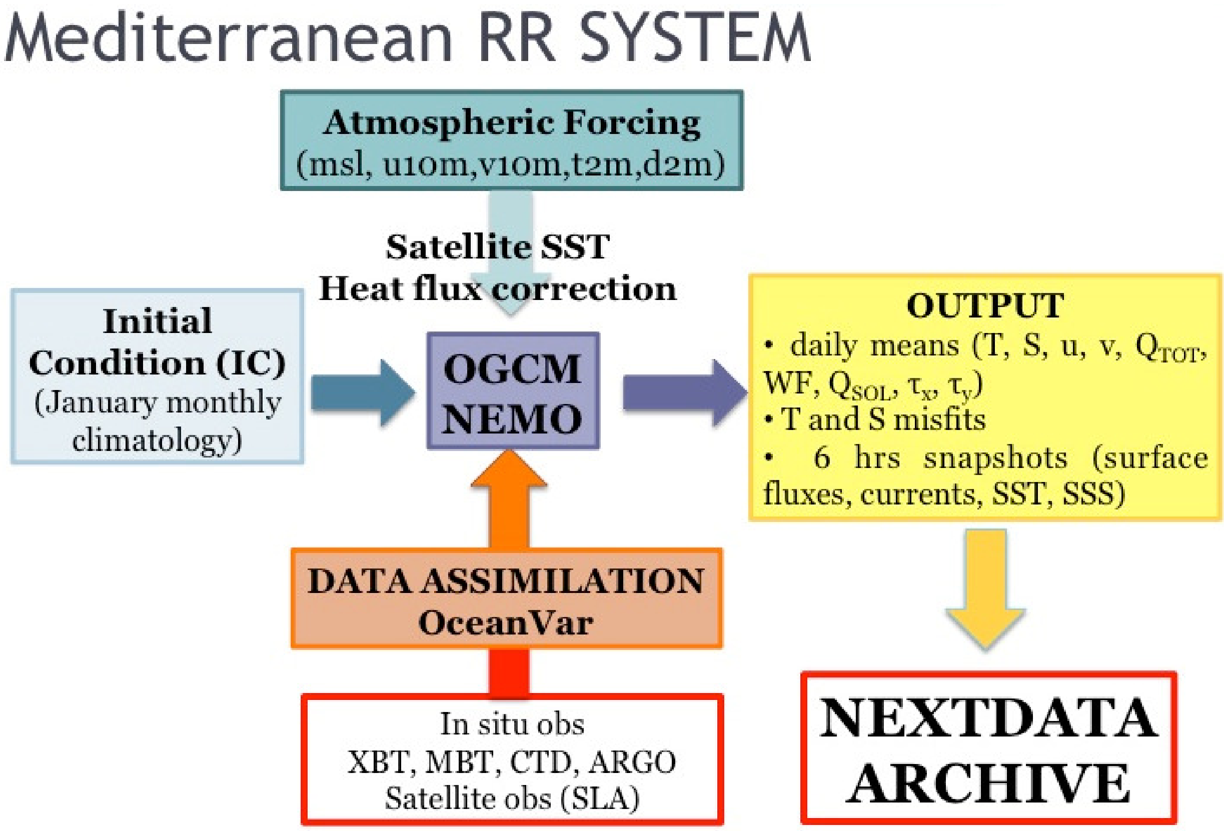


Figure 1. Schematic of the Mediterranean Reconstruction/Reanalysis (RR) system and its components.

	MedReanV2 (1985-2007)	MedReanV4(1985-2010)	RR100 (1912-2011)
SST	SST reconstruction (1985-2007) *Marullo et al.2007*	SST reconstruction(1985-2007) *Marullo et al.2007* MyOcean data (2008-2010)	**HadISST1** (1912-1985) **SST reconstruction** (1985-2007) *Marullo et al.2007* **MyOcean data** (2008-2011)
SLA	ERS1, ERS2, EnviSat, TOPEX/ Poseidon, Jason1 (Pujol and Larnicol 2005)	MyOcean multisensor "UPD" data reprocessed in 2010	MyOcean multisensor "UPD" data reprocessed in 2010
XBT	<1999 MEDATLAS, MATER Ship of Opportunity *(Manzella 2007)* 2000-2007 MFS observations	SeaDataNet MFS and MyOcean obs. (1985-2010)	SeaDataNet MFS and MyOcean obs. (1912-2011)
ARGO	MedArgo Program (2001-2007) *Poulain et al. 2007*	All observations (2001-2010)	All observations (2001-2011)
CTD	MEDATLAS, MATER (<1999) MFS observations (2000-2007)	SeaDataNet MFS and MyOcean obs (1985-2010)	SeaDataNet MFS and MyOcean obs. (1912-2011)

Tab. 1 Historical marine datasets used in the three reconstruction/reanalysis systems at INGV: 1) MedReanV2 already available from *Adani et al. (2011)* within the framework of CIRCE Project; 2) MedReanV4 in production within the framework of MyOcean Project; 3) RR100 is the NEXTDATA developing system for the Mediterranean Sea.

Remote Sensing Observations

1. SLA

Altimetry data come from **MyOcean** (implementation project of the GMES Marine Core Service) **Sea Level TAC** (Thematic Assembly Centre). They are delayed time (DT) SLA data "upd" version released in March 2010, the latest version of AVISO reprocessed data present in the ftp site. The Delayed Time component of SSALTO/DUACS system is responsible for the production of processed Jason-1, Jason-2, T/P, Envisat, GFO, ERS1/2 data in order to provide a homogeneous, inter-calibrated and highly accurate long time series of SLA. Figure 1 of deliverable D2.2.1 shows the time period of satellite altimetry monitoring and the various sensors coverage. RR system will assimilate these data starting from 1992.

2. Satellite SST

Satellite SST is used to correct interactively the computed heat flux at air-sea interface as schematized in Figure 1. The more recent SST dataset is a time concatenation of many SST products characterized by horizontal maps already optimally interpolated onto the RR model grid at 1/16th of a degree:

1. **reprocessed data** (1985-July 2008) of the recent AVHRR Pathfinder SST (*Marullo et al., 2007*);
2. **MyOcean SST-TAC** (Thematic Assembly Centre) data (Jul2008-Aug2010);

For the time period preceding 1985 we will consider the Met Office Hadley Centre's sea SST data set, **HadISST1**, a unique combination of monthly globally-complete fields of SST on a 1 degree latitude-longitude grid starting from 1870. This choice has been done consistently with the idea to use AMIP (Atmospheric Model Intercomparison Project - *Gates, 1992*) type of atmospheric data as surface forcing for the pre-ERA40 period starting from mid-1957 (see deliverable D.1.3.2 for further details). AMIP type of data (*Cherchi and Navarra, 2007*) are available starting from 1900 up to 2003 and were created through a set of experiments performed with the ECHAM4 atmospheric global circulation model (AGCM) on a T126 grid (1.125deg) forced by the HadISST1 dataset and saved twice a day. In particular 7 experiments were archived (WP2.2) with similar characteristics but they must be analyzed and compared to the available ECMWF reanalysis products (ERA-40 and ERA-Interim) to choose the best one to use in the RR.

Historical In Situ Temperature and Salinity Observations

The first year work focused on observational data analysis. The in situ temperature and salinity profiles considered for the RR production belong from different instrumental data type: CTDs, XBTs, MBTs, bottles, ARGO. They are summarized in Table 1 together with the previous reanalysis products developed at INGV: **MedReanV2** from *Adani et al. (2011)* produced within the framework of CIRCE Project (http://www.circeproject.eu/) and **MedReanV4** in production within the framework of MyOcean Project.

In situ data sets have been collected from European Marine databases (SeaDataNet, GMES) and have been archived as described in deliverable D2.2.1 in a specific format to be assimilated in the RR. They were downloaded from three main sources:

1. **SeaDataNet** (SDN hereafter) European infrastructure (DG-Research – FP6);
2. **MEDAR-MEDATLAS** dataset covering the period 1985-1999 (*Maillard et al. 2005*);
3. **MFS** (Mediterranean Forecasting System) and **MyOcean** (implementation project of the GMES Marine Core Service). In situ TAC (Thematic Assembly Centre) operational observations.

The merging of these data sets was necessary due to missing data within the SDN infrastructure. **SDN historical database** gathers in situ data from about forty NODCs (National Ocean Data Center) and its data population started and progressively increased on the first phase project implementation. The second phase of SDN project is now devoted to the assessment of the quality of the database content and the duplicate elimination.
Potential **duplicates** were thus identified and excluded from successive usage and analysis. Looking at the annual data distribution of the number of temperature and salinity observations within SDN database (see Figure 2 of D2.2.1) from 1990 to 2012, it is evident the decrease of the number of observations for the recent years due to a time lag between the sampling and the insertion of the data inside the SDN infrastructure, which is a common characteristic of historical databases. This required the use of MFS and MyOcean in situ TAC operational observations to integrate the SDN data set in the recent period. We intend for **MFS operational observations**, near real time (NRT) observations collected in the Mediterranean Sea within different precursor projects spanning a time period from 1999 to April 2009 when MyOcean Project started. Precursor European projects are:

- **MFSPP** (Mediterranean ocean Forecasting System Pilot Project) **1998-2001** EU-MAST project MA 53-CT98-0171;
- **MFSTEP** (Mediterranean ocean Forecasting System Towards Environmental Prediction) **2003-2005** DG-Research – FP5 EU Contract Number EVK3-CT-2002-00075;
- **MERSEA** (Marine Environment and Security for the European) **Strand 1** 2001-2003;
- **MERSEA Integrated Project** 2004-2008.

Another check has been implemented to verify the presence of **MEDAR-MEDATLAS** data set (assimilated in **MedReanV2** *Adani et al., 2011)* within SDN infrastructure. Missing data have been integrated to obtain the most extensive dataset.
Figures 3 and 4 of D2.2.1 show the annual distribution of the number of temperature and salinity profiles for the time period 1959-1989 and 1911-1958. The time period 1911-1946 is characterized by very few observations and in many years there are no observations at all. The availability of is situ observations starts and systematically increases from 1946.

Quality Control Procedure

SDN data are quality controlled and managed at distributed data centers (NODCs), interconnected by the SDN infrastructure and accessible through an integrated portal. The data sets come from various sources and time periods. This needed strong requirements towards ensuring quality, elimination of duplicate data and overall coherence of the integrated data set. This has been achieved through the definition of accurate metadata directories, common standards for vocabularies, metadata formats, data formats, quality control methods and quality flags (Tab.2).

In situ observations were selected from the entire downloaded data set through the **general quality flag scheme** (see Tab.2) applied consistently by SDN, MyOcean and MEDAR-MEDATLAS **automated basic quality checks**. Ocean Data View (ODV) software http://odv.awi.de/ has been used to perform the QC analysis. It can easily manage big amount of data and it has been developed to be compliant with both SDN and MEDAR-MEDATLAS data format.

In particular we selected the profiles with general flags (Tab.2) equal to 1 related to their right positioning:
(I) date and time of an observation has to be valid:

- Year 4 digits – this can be tuned according to the data
- Month between 1 and 12
- Day in range expected for month
- Hour between 0 and 23
- Minute between 0 and 59

(II) latitude and **longitude** have to be valid:

- Latitude in range -90 to 90
- Longitude in range -180 to 180

(III) position must not be on land

- Observation latitude and longitude located in ocean.

Quality flags associated to **depth** equal to 0, 1 and 2 (refer to Tab.2) were selected as shown in Figure 2(a) in a screenshot of ODV software. These flags relates to:
(VI) deepest pressure

- The profile does not contain pressures higher than the highest value expected (or no data points below bottom depth)
- Check that pressure is monotonically increasing

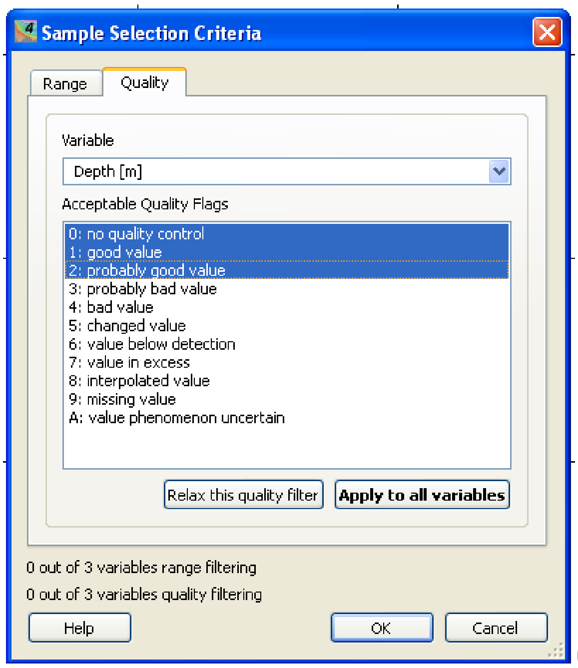
Then further quality control has been carried out on the data set through **visual inspection** using ODV. There is often a subjective element in this process. A **gross range check** (Figure 2b) has been applied to avoid unrealistic temperature (4 <T<30C°) and salinity (15<S<40psu) values. Duplicate check has been applied again at this stage (Figure 2c) considering specific tolerance intervals on time (0.1 days), longitude (0.02deg) and latitude (0.02deg). Then the ODV built in **spike detection** analysis has been used before saving the full data set with the relative quality flags in a spreadsheet file compatible with DIVA software http://modb.oce.ulg.ac.be/mediawiki/index.php/DIVA utilized to compute temperature and salinity gridded fields, as it will be better detailed in deliverable D1.3.2 on RR system configuration.

Observed profiles needed also to be written on ASCII files to be read from the OceanVar data assimilation scheme (D2.2.1). A further automated QC is performed during this phase considering the quality flags associated to temperature, salinity and depth at each vertical level of the observed profiles. Only the profiles having less than the 25% of bad data are assimilated in the reanalysis. Bad data are intended data that have not all three flags equal to one.

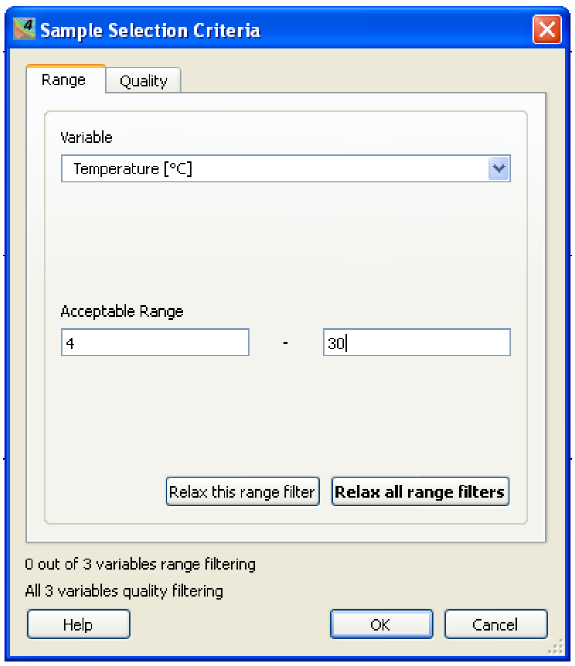
In Figure 3 is shown the different data set assimilated by **MedReanV2** reanalysis (*Adani et al. 2011*) and by **MedReanV4** in the time period that goes from January 1985 to October 1987 in the Northern Ionian Sea. It can be noticed that in MedReanV2 were not assimilated POEM 1987 survey observations (*Robinson et al. 1991*) not included in MEDAR-MEDATLAS data set.

Key	Entry Term	Abbreviated term	Term definition
0	no quality control	none	No quality control procedures have been applied to the data value. This is the initial status for all data values entering the working archive.
1	good value	good	Good quality data value that is not part of any identified malfunction and has been verified as consistent with real phenomena during the quality control process.
2	probably good value	probably_good	Data value that is probably consistent with real phenomena but this is unconfirmed or data value forming part of a malfunction that is considered too small to affect the overall quality of the data object of which it is a part.
3	probably bad value	probably_bad	Data value recognised as unusual during quality control that forms part of a feature that is probably inconsistent with real phenomena.
4	bad value	bad	An obviously erroneous data value.
5	changed value	changed	Data value adjusted during quality control. Best practice strongly recommends that the value before the change be preserved in the data or its accompanying metadata.
6	value below detection	BD	The level of the measured phenomenon was too small to be quantified by the technique employed to measure it. The accompanying value is the detection limit for the technique or zero if that value is unknown.
7	value in excess	excess	The level of the measured phenomenon was too large to be quantified by the technique employed to measure it. The accompanying value is the measurement limit for the technique.
8	interpolated value	interpolated	This value has been derived by interpolation from other values in the data object.
9	missing value	missing	The data value is missing. Any accompanying value will be a magic number representing absent data.
A	value phenomenon uncertain	ID_uncertain	There is uncertainty in the description of the measured phenomenon associated with the value such as chemical species or biological entity.

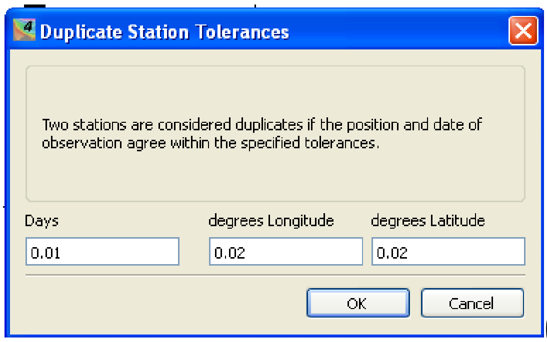
Tab. 2. SeaDataNet quality flags (L201)

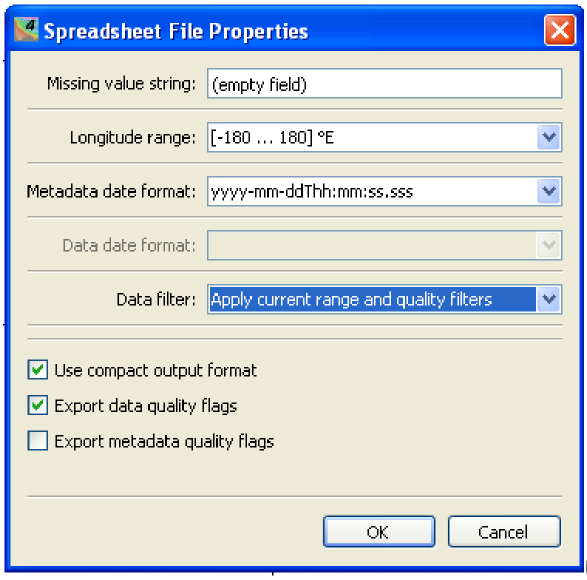


(a)



(b)



(c)


(d)

Figure 2. Ocean Data View (ODV) screen shoots of different quality control (QC) phases: a) quality flag selection; b) range check; c) duplicate detection; d) data export.

Conclusions and Future Work

The necessary observational data sets have been collected for the RR production during the first year of NEXTDATA Project and the basic QC procedures have been performed in order to assimilate data into the RR system and compute the gridded fields climatologies for model initialization and successive validation.

Future work will be devoted to:

- The refinement of QC and data assimilation pre-processing procedures in order to continuously ameliorate the future RR product quality;
- A constant data set update;
- The discovery of new independent data sets to be used for the quality assessment of the RR product.

A constant archived data set upgrade is necessary since SDN historical database is continuously refined as additional quality checks are undertaken.

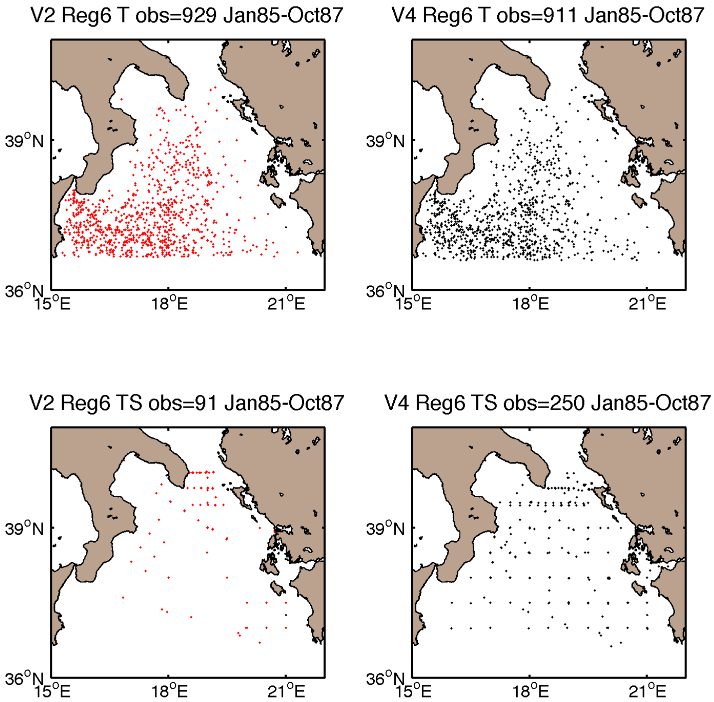


Figure 3. Data distribution map of XBT and CTD assimilated in the period Jan1985-Oct1987 from: (left top and bottom) MedReanV2; (right top and bottom) MedReanV4.

References

Adani M., Dobricic, S. Pinardi N., 2011: Quality Assessment of a 1985–2007 Mediterranean Sea Reanalysis. J. Atmos. Oceanic Technol., **28**, 569–589. doi: http://dx.doi.org/10.1175/2010JTECHO798.1

Cherchi, A. and A. Navarra, 2007: Sensitivity of the Asian summer monsoon to the horizontal resolution: differences between AMIP-type and coupled model experiments, Climate Dynamics, 28, 2, 273-290, doi:10.1007/s00382-006-0183-z

Dobricic S., Pinardi N., 2008: An oceanographic three-dimensional variational data assimilation scheme, Ocean Modelling, Volume 22, Issues 3–4, , Pages 89-105, ISSN 1463-5003, 10.1016/j.ocemod.2008.01.004.

Gates, WL, 1992. AMIP: the atmosphere model intercomparison project. Bulletin of American Meteorological Society 73, 1962–1970.

Maillard, C., and Coauthors, 2005: MEDAR/MEDATLAS 1998-2001: A Mediterranean and Black Sea oceanographic data base and network. Boll. Geofis. Teor. Appl., 46, 329–344.

Madec, G., P. Delecluse, M. Imbard, and C. Levy, 1998: OPA 8.1 ocean general circulation model reference manual. Institut Pierre-Simon Laplace, Note du Pole de Modelisazion, No. 11, 91 pp.

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

Rayner, N. A.; Parker, D. E.; Horton, E. B.; Folland, C. K.; Alexander, L. V.; Rowell, D. P.; Kent, E. C.; Kaplan, A. (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century J. Geophys. Res.Vol. 108, No. D14, 4407 10.1029/2002JD002670

A.R. Robinson, M. Golnaraghi, W.G. Leslie, A. Artegiani, A. Hecht, E. Lazzoni, A. Michelato, E. Sansone, A. Theocharis, Ü. Ünlüata, The eastern Mediterranean general circulation: features, structure and variability, Dynamics of Atmospheres and Oceans, Volume 15, Issues 3–5, April 1991, Pages 215-240, ISSN 0377-0265, 10.1016/0377-0265(91)90021-7.

Tonani, M., Pinardi, N., Dobricic, S., Pujol, I., and Fratianni, C., 2008: A high-resolution free-surface model of the Mediterranean Sea, Ocean Sci., 4, 1-14, doi:10.5194/os-4-1-2008.