Progress on Quality Control in SADCO
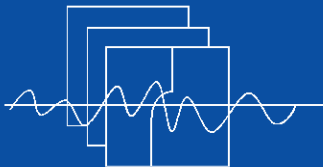
SADCO has amended the data base structure to accommodate Quality Control flags. Already flagged data can now be loaded along with its flags, while a retroactive screening and flagging of all data loaded previously is foreseen.

1. VOS data

A few years ago SADCO analysed the spread of values of parameters contained in the VOS (Voluntary Observing Ships) observations (surface winds, air temperature, etc). Based on these values from our (local) target area a set of envelopes were derived within which values should be located. These envelopes are now routinely applied to filter all incoming VOS data.

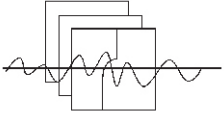## 2. Present marine (hydrographic stations) data checks

a. Present checks during loading

- **Position.** There are no independent sources against which the position of a hydrographic station can be checked (log books are not available). The only "check" is to calculate the speed with which the vessel would have had to travel between successive stations, and inspect unreasonably high values (the nominal threshold is 20 knots). If the speed anomaly can be corrected by judiciously amending the time of a station, this option is preferred. A plot is also made of cruise tracks, and an anomalously located station is often quite evident as a positional spike. In a previous Newsletter we reported on the on-screen identification of such anomalies, although any corrections still need to be handled off-line. A clue to the correct position can often be found if the spatial pattern of the stations are regular (straight line, or regular coastal transects).

- **Date and time** are initially checked for validity only (limits of the year, month, day of the month, time) and incorrect entries are often also evident in the speed checks (see above).

- **Parameter profiles.** Parameter envelopes (temperature, salinity oxygen) have been extracted from the World Ocean Database, and incoming data is screened using these envelopes. Up to now, significant outliers have been eliminated. The recent data obtained from the World Ocean Database 2005 were not subjected to this checking, since the data had already been screened and supplied with Quality flags. The data was loaded along with these flags.

3. Planned QC checking on hydrographic stations

a. Installation of quality flags.

SADCO's understanding has always been that the data donor is the primary controller of the data quality. It may have been that, many years ago, data submitted to SADCO was completely without errors, but it is suspected that, in the absence of detailed checking, data was loaded whether they had errors or not. It is also assumed that, because the data mostly came from marine organisations close by, individual data providers could be contacted for clarification if ever erroneous data was followed up. Given the large amounts of data presently handled by virtually all marine data centres, it has been shown that a better way of dealing with quality issues is to tag data that do not conform to agreed checks. SADCO has now, for the first time, amended the data bases to provide the space for quality flags, an essential first step in the large scale quality control process. The definition of the flags is the same as those employed by NODC (Washington) and address the known errors that appear in profile data.

b. Example of flags.

Observations that lie outside the envelopes are flagged individually if they fail the monthly, seasonal or annual envelope check. SADCO is presently employing only the annual envelope check, since the data base for the shorter-term checks (seasonal, monthly) is so small that the envelopes become somewhat "jagged" and could lead to spurious anomalies.

c. Loading of data that already contain flags.

With the installation of the space for quality control flags, incoming data that has already been checked and flagged (according to the same criteria) where required will be transferred together with their QC flags.

Because the flags are only installed now, while large amounts of data have been loaded in SADCO in the past (> 200 000 stations), a retroactive screening of the data needs to be undertaken. The modus operandi will be that the QC programme will systematically identify anomalies and install the flags on ALL the data, thus supplying the flags in one process.

4. Future checking

It is foreseen that, once the tasks indicated above have been completed, incoming data will be checked and QC flagged during the load process.

5. Outcome

Eventually, all data will contain the necessary flags and, for the first time in its history, SADCO will be able to differentiate according to accepted norms on the quality of the data.

It then remains to amend the extraction and product routines to the flagged data, possibly with choices on whether a certain product/extraction should be based on ALL data or only data that does not contain certain flags.
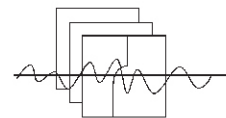
Summary review of SADCO Activities, 2006/7

In a very busy 2006/7 data loading still comprised the largest activity of the data centre. A part of this activity was the loading of about 19 000 stations from the World Ocean Database 2005 (historic data collected in SADCO's target area by foreign vessels mostly before 1980). More data from this source will be loaded in 2007/8.

The following information has been extracted from the 2006/7 SADCO Annual Report submitted to the Steering Committee in May 2007.

1. SADCO Memorandum of Understanding (MoU)

The draft MoU, elucidating the scope of activities and responsibilities of the data centre and the Steering Committee, as well as the support of the various organisations for SADCO) was tabled at the May 2005 meeting of the Steering Committee and approved to be signed by all the stakeholders. By May 2006 the MoU had been signed by 4 of the 7 organisations, and by May 2007 it had been signed by 6 of the 7.

2. Interaction with SAEON Offshore Node

There have been preliminary discussions between SADCO and SAEON (South African Environmental Observation Network) about the foreseen activities of the Offshore Node, and the interaction between SADCO and the Offshore Node. These discussions are being followed up with Dr Juliet Hermes, manager of the SAEON Offshore Node.

3. AFROBIS

The RUTGERS-CSIR seed funding to establish and manage AFROBIS (African node of the Ocean Biogeographic Information System) ran until 30 November 2006 (see the review of overall achievements in the SADCO Newsletter of December 2006). Progress on AFROBIS was presented at all SADCO meetings, and the Steering Committee expressed its satisfaction with the achievements. New funding streams are being pursued to establish long-term sustainability for the database.
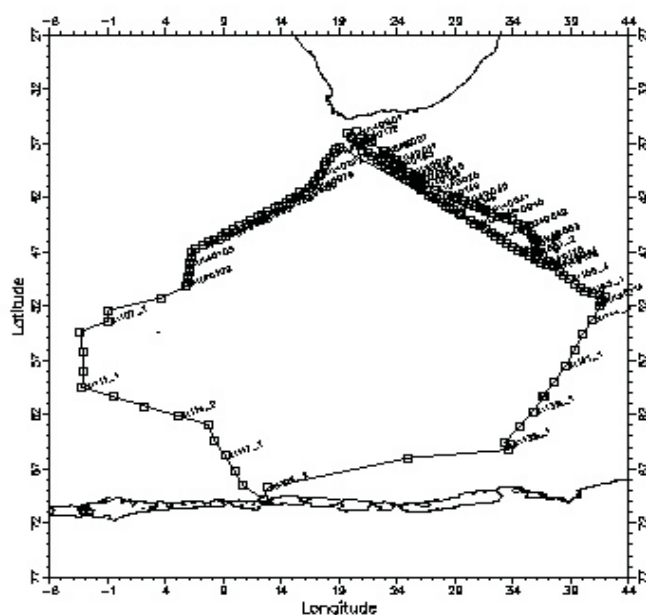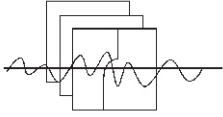

Fig. 1 Example of a cruise track chart of cruise ANT XI Leg 4, Mar-May 1994, obtained from the Alfred Wegener Institute (AWI). (See Item 5, p.4).

Summary review of
SADCO Activities, 2006/7 *(continued..)*

4. Enhancement of marine loading program

In view of the planned quality control procedures in SADCO, as well as the loading of data from WOD2005 (see 6. below) the marine database was modified to include quality control flags for the first time (for more details see Article on Quality Control on p.1 of the Newsletter).

The actual parameter checking will be dealt with through a separate programme, which will also be the cornerstone of the marine data cleanup task (retroactively posting quality flags on the data), which has not been started yet.

5. CTD data from individual organisations

A summary of the CTD and XBT data loaded during 2006/7 is provided in Table 1, and an example of a cruise track is shown in Fig. 1.

6. Data from the *World Ocean Database 2005*

The World Ocean Database 2005 (WOD2005) also contains data collated as part of the Global Ocean Data Archiving and Rescue (GODAR) project run by the World Ocean Data Centre in Washington. In 1999 SADCO loaded approximately 50 000 stations gleaned from the results of this project.

Data sets that were added to the WOD after the previous SADCO loading, could be equally useful to SADCO. This data is contained in three subsets, namely OSD (bottle data, hydrosondes), CTD, and XBT/MBT. Thus far, data from the OSD subset was identified, extracted and loaded.

A total of 19 066 stations from 265 cruises were loaded. Fig. 2 provides examples of the track charts of some of the cruises.

The data from the CTD and XBT/MBT sets will be loaded in 2007/8.

Table 1. Summarised loading status of CTD and other profile data

Origin	Data	Number of stations
UCT	GoodHope Feb/Mar 2004 XBT	185
	GoodHope Nov 2004 XBT	118
Japan	Profiles	50
Germany (AWI)	XBT	2572
	CTD	806
	Total	3731
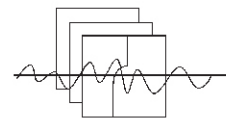
Summary review of
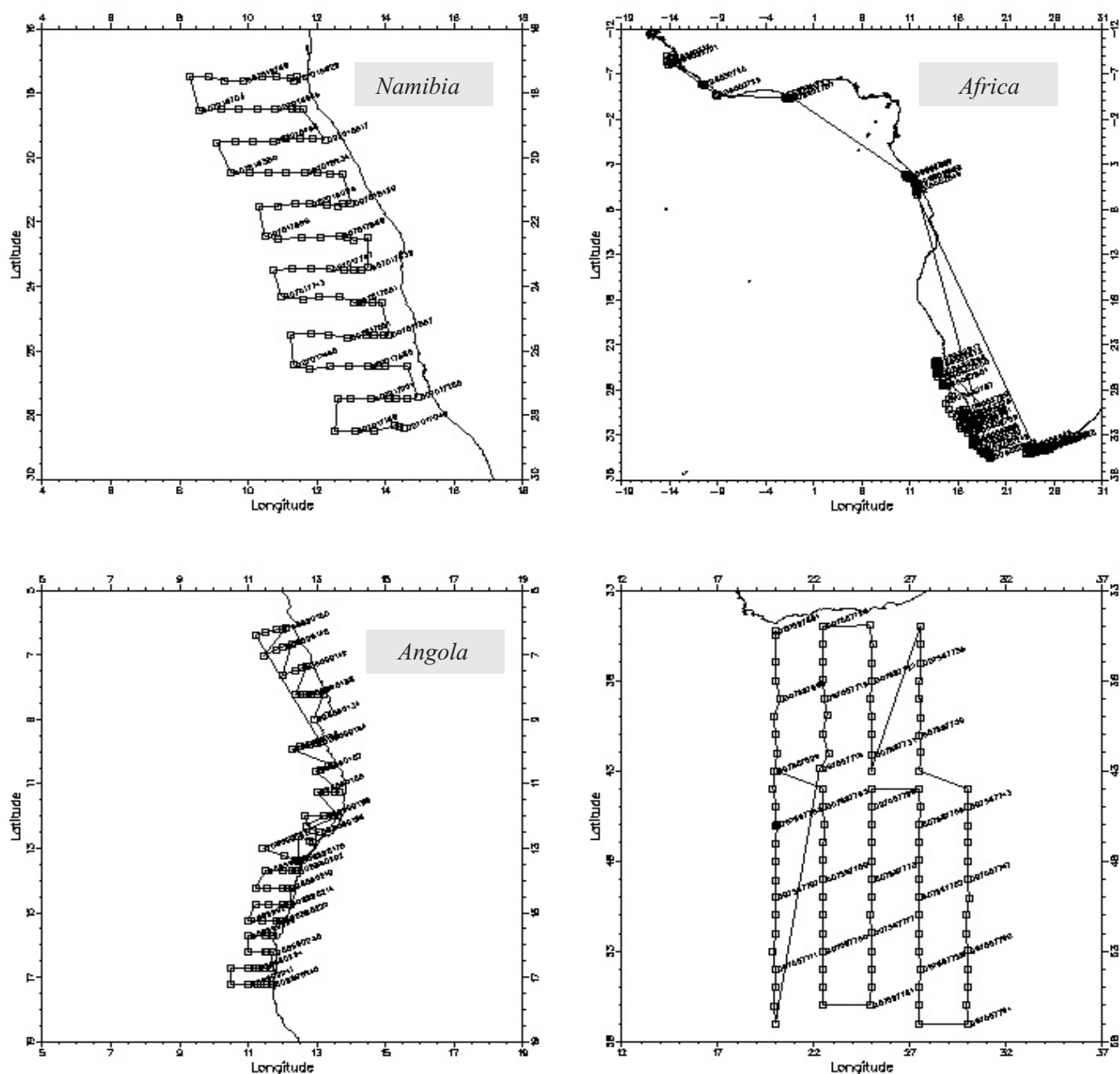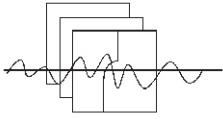# SADCO Activities, 2006/7 *(continued..)*

Fig. 2 Examples of cruise tracks from the WOD2005 data loaded. **Top left:** Nekton cruise, Jul-Aug 1985; **Top right:** Unknown vessel, Feb-Jul 1971; **Bottom left:** Alferas cruise, Jun-Jul 1978; **Bottom right:** Professor Zubov, Dec 78 – Feb 79.

Summary review of
SADCO Activities, 2006/7 *(continued..)*

7. ARGO float data

SADCO has also started to scout for data from the ARGO floats, and the data will be loaded in 2007/8.

8. MCM moored ADCP data

The set of moored ADCP (Acoustic Doppler Current Profile) data from Tsitsikamma that had been submitted to SADCO by Mike Roberts (MCM) has been loaded. The data covers the period December 2003 to November 2005.

9. SAWS VOS data

VOS (Voluntary Observing Ships) data loaded during the year is indicated in Table 2. This data is prepared and made available by the South African Weather Service, and regularly downloaded from Pretoria. The loading process is a fluent operation, with the minimum of human intervention. Data is checked according to the QC screening procedures created previously. (See p.1 of this Newsletter).

10. Weather station data

- AWS (automatic weather station) data collected on Roman Rock (Institute for Maritime Technology) submitted to SADCO last year has now been loaded. The data covers 4 years, from 1 January 2003 to 31 December 2006, with recording every 10 minutes.
- The coastal weather data received from the South African Weather Service last year has been loaded. The data is recorded at hourly intervals, and covers the period 1995 to 2005.
- Although no AWS data has been submitted by MCM during the year, Table 3 provides a listing of all data loaded so far.

Table 2. Number of VOS data records loaded

Month	Main base*	Archive*
Apr-06	1869	2099
May-06	1538	1472
Jun-06	1138	1196
Jul-06	2183	2222
Aug-06	1512	1477
Sep-06	1407	1583
Oct-06	1541	1639
Nov-06	1331	1399
Dec-06	1890	2367
Jan-07	1440	1472
Feb-07	1890	1512
Mar-07	1512	1537
TOTAL	**19251**	**19975**

* "Main base" refers to the area between 0 and 50º E, and after 1960. "Archive" refers to the rest of the data (time wise and space wise)

Table 3. Automatic Weather Station data from MCM

Location	Period
Danger Point	Jan – Jun 1997; Jan – Dec 1998; Dec 2001 – Dec 2002
Cape Columbine	Oct – Dec 85; Dec 93 - Dec 2002
Port Nolloth	4 Dec 2001 – 3 Dec 2002
Diaz Point	Aug – Nov 89
Hondeklip Bay	Sep 82 – Jun 84; Nov 88 – Feb 91; Nov 91 – Feb 95
Kleinzee	Nov 93 – Jan 94
Nuwedam	Aug 85 – Feb 86
Olifantsbos	Nov 89 – Jun 97
St Francis Bay	Aug – Oct 94
Stompneus	Nov 83
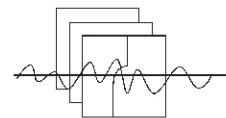
Summary review of
SADCO Activities, 2006/7 *(continued..)*

11. CSIR chemical data

The marine section of the CSIR in Durban has been collecting chemical data off the KwaZulu-Natal coast for many years. These surveys have been mostly aimed at the monitoring of point sources of marine effluent disposal in the region.

The database was amended to accommodate such data and appropriate software was written for the loading.

For the Richards Bay pipeline water chemistry data from August 1987 to October 2004, and sediment chemistry from May 1987 to October 2004 has been reformatted and loaded. The full set for Richards Bay water chemistry goes back to February 1977, and for sediment chemistry to May 1973.

For the two Durban pipelines, water and sediment chemistry from May 2001 to October 2004 has been loaded.

12. DeBeers Marine current meter data

A visit was paid to DeBeers Marine, where Lesley Roos kindly provided SADCO with current meter data collected off Oranjemund. This data will be loaded in 2007/8.

13. Request handling

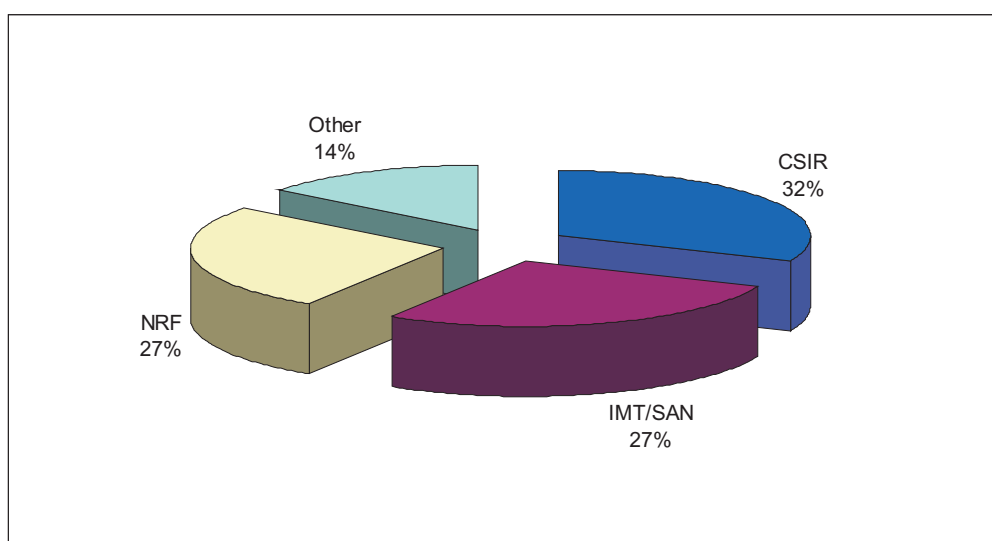The diagram below (Fig, 3) shows the distribution of the origin of the requests submitted to SADCO.
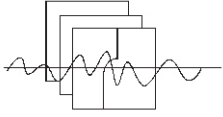

Fig. 3.
Distribution of requests
handled by SADCO

SADCO Work plan 2007/8

At the May meetings of the SADCO Steering Committee every year items are tabled and discussed that should form part of the activities of the Data Centre for the coming year. These activities exclude the day-to-day and routine activities to ensure the viability and performance of the data centre. Below is a list of activities decided in May 2007 that will be addressed in the period up to March 2008.

It should be noted that there is a degree of "overestimation" of the number of activities that can be achieved. This overestimate arises from the fact that the full duration of each task is largely unknown at the start (some of the tasks will take more than a year to complete). There is also reference to data sets that should be loaded but that have not been received yet.

- Finalise MoU
- Finalise SADCO Strategy 2005-2010
- Software for quality control of parameters
- Marine database cleanup (retroactive)
- Data transfer to the World Data Centre
- Data loading:

  - Moored ADCP data (Namibia)
  - Moored ADCP data (AWI)
  - Current meter data Oranjemund (DeBeersMarine)
  - Moored ADCP data from MCM (*if submitted*)
  - Finalise loading data from WOD2005 (CTD, XBT)
  - ARGO float data
  - Other surface drifters
  - DEIMEC CTD data (UCT)
  - Continuous and discrete CTD data from MCM (*if submitted*)

  - AWS data from SAWS (Robben Island, Gough Island)
  - Mozambique/NANSEN wind data
  - AWS data from Roman Rock (IMT) (ongoing, *if submitted*)
  - AWS data from MCM (ongoing, *if submitted*)

  - Additional chemical data (CSIR) (*if submitted*)

  - CLIWOC data (1750 – 1850)

Because each task has its own life cycle, the order in which tasks are started and completed is a function of the duration of each task, its complexity and associated preparations, the availability of the data donor for queries, etc.