Vol 18 No 3 - August 2007

SADCO

SADSO

Loading data from the World Ocean Database 2005

In 1999 SADCO loaded data from the World Ocean Database (WOD98) (supplied by the World Data Centre, WDC, in Washington), and a chart with all the stations was provided in a Newsletter in 2000. Approximately 50 000 stations were loaded at that time – the largest single load effort ever undertaken by SADCO. Although updated versions of WOD have been produced since then, SADCO tentatively decided to wait for a few years before tackling a new loading session for this data.

The WOD2005 was searched for data within SADCO's target area, and the following were extracted and loaded:

OSD: 17 351 stations*. These are bottle data, TS profilers, etc.

CTD: 1287 stations

XBT: 8691 (expendable bathythermograph)

MBT: 8904 (mechanical bathythermograph)

These stations have been plotted in Fig. 2. Where a cruise was located mainly inside SADCO's target area but also extended outside of the area, those stations were also loaded. These latter stations have not been plotted in the Figure.

The following aspects are noticed from the graph:

- As before, the station distribution is somewhat dominated by the traditional shipping lanes, where most of the XBTs were deployed.
- There is a noticeable density of stations off Namibia, and also off the Mozambique coast. For interest, the area off Namibia has been enlarged in Fig. 1.

We want to thank the WDC for compiling the WOD and making the data available.

* In the previous Newsletter a total of 19 066 was reported. A rescreening of the data rejected about 1 600 stations on the basis of doubtful positions (a fatal flaw that will prevent the data from passing the quality control process – see article this Newsletter) or the bulk of the cruise proving to be outside the SADCO target area.

Southern African Data Centre
for Oceanography
P O Box 320, Stellenbosch 7599
South Africa

Manager: Marten Gründlingh

Email: mgrundli@csir.co.za

Website: http://sadco.csir.co.za/

SADCO is sponsored by ...

Department of Environmental Affairs & Tourism
SA Navy
CSIR
NRF (SA Universities)
Namibian Ministry for Fisheries & Marine Resources

ISSN 1017-2351

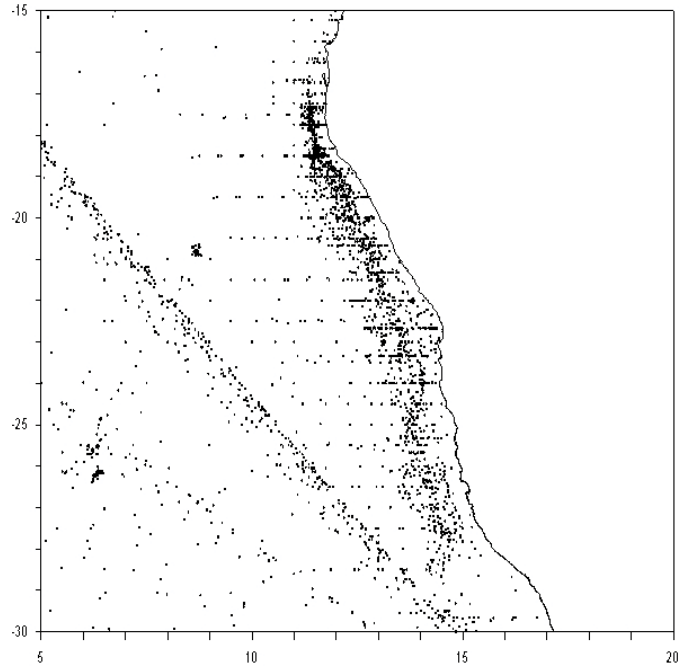


Fig. 1 Detail of stations off Namibia

Quality control on vertical hydrographic profiles

Ursula von St Ange

SADCO has completed the quality control measures for its hydrographic data. The data is screened on station level, profile level, as well as individual (subsurface) observation level. The checks are a subset of those designed by the World Ocean Database. The finalisation of this checking process marks a significant milestone in SADCO's QC process, and allows all data to be classified retroactively.

Previous newsletters have indicated the various quality control (QC) measures employed by SADCO to assess the correctness of data.

In general, data centres do not correct erroneous data without consulting with the data provider. However, more and more data (loaded by SADCO) do not originate from local, "consultable" sources, but have been collected by overseas organisations (that may not even exist any more). Such data is often relayed via a global data centre (e.g. through GODAR – Global Ocean Data Archiving and Rescue project - of the World Data Centre).

Data centres have therefore created a system of flagging, whereby questionable data is indexed according to its agreement with accepted values, or a range of values, without removing the data itself.

SADCO recently decided to install quality flags on all its historic subsurface data (about 250 000 hydrographic stations). To conform with international checks, the QC flags of the World Ocean Database (WOD) were adopted, so that data obtained from the World Data Centre (WDC) could be be integrated almost seamlessly into the existing SADCO data base. In addition, the process will be applied retroactively to historic data in batch mode.

Description of the QC flags

The QC process involves 3 groups of checks, namely checks of the **station** information, checks of the **profiles** and checks of the **subsurface observations**. No flags are installed for the station checks (stations are either accepted – sometimes with small corrections - or rejected). For the **profile** and **observation** checks, flags are allocated per parameter (e.g. temperature, salinity, nutrients,...) and the flags installed by SADCO are a

Quality control on vertical hydrographic profiles *(continued..)*

subset of the WOD flags plus a spike check (see Table 3.) The variables that are checked are: depth, temperature, salinity, oxygen, phosphate, silicate, nitrate, pH, dissolved inorganic carbon (DIC) and total chlorophyll.

1. Station checks

a) Position / date / time

- Checks are made for invalid latitudes, longitudes, dates and times (i.e. values that lie outside certain limits) during the loading process. Stations of which the position/date/time are invalid are not loaded.
- If the speed between adjoining stations is unrealistic, the date/time and/or the position may be wrong. This is checked manually, and if the corrections are obvious they are applied, otherwise a station is not loaded.

b) Duplicate stations

Before each station is loaded, a check is done on the database for all stations that have been collected by the same ship temporarily within 10 minutes and spatially within 2 nautical miles of the new station. The data type (CTD, XBT, etc.) and the number of depths are also compared. If everything confirms that the incoming station is a duplicate, it is rejected.

2. Profile flags

a) Standard deviation (envelope)

As the statistical data for the *seasonal* and *monthly* data was too sparse, SADCO decided to only implement the annual standard deviation check, using "envelopes" provided by WOD for 5ºx5º blocks. These envelopes are determined from an average value PLUS and MINUS a number of standard deviations. For "coastal" stations, 5 standard deviations are used, for "near coastal" 4, and for "open ocean" 3.

If e.g. a temperature profile contains two or more values that lie outside the envelope, the profile is flagged.

b) Density inversions

The determination of the density inversion is described in 3(b). If two or more such density inversions occur on one station, the temperature and salinity profiles are both flagged.

c) Spike and additional gradient checks

SADCO does an additional test for spikes and gradients in the temperature and salinity profiles, using IOC (Intergovernmental Oceanographic Commission) algorithms. Checks are separately done for top and bottom spikes, spikes in the rest of the profile, as well as excessive gradients.

*Check for the **top spike***

if $v_{lower\ limit} < (v1 - v2) < v_{upper\ limit}$, is untrue, v2 is flagged

*Check for **bottom** spike*

if $v_{lower\ limit} < (v2 - v1) < v_{upper\ limit}$, is untrue, v2 is flagged

*Check for a **spike in the rest of the profile***

if $|(v2 - (v1 + v3))/2| - |(v1 - v3)/2| > v_{threshold}$,
then v2 is flagged

*Check for **excessive gradient***

if $|(v2 - (v1 + v3))/2| > v_{grad}$, then v2 is flagged

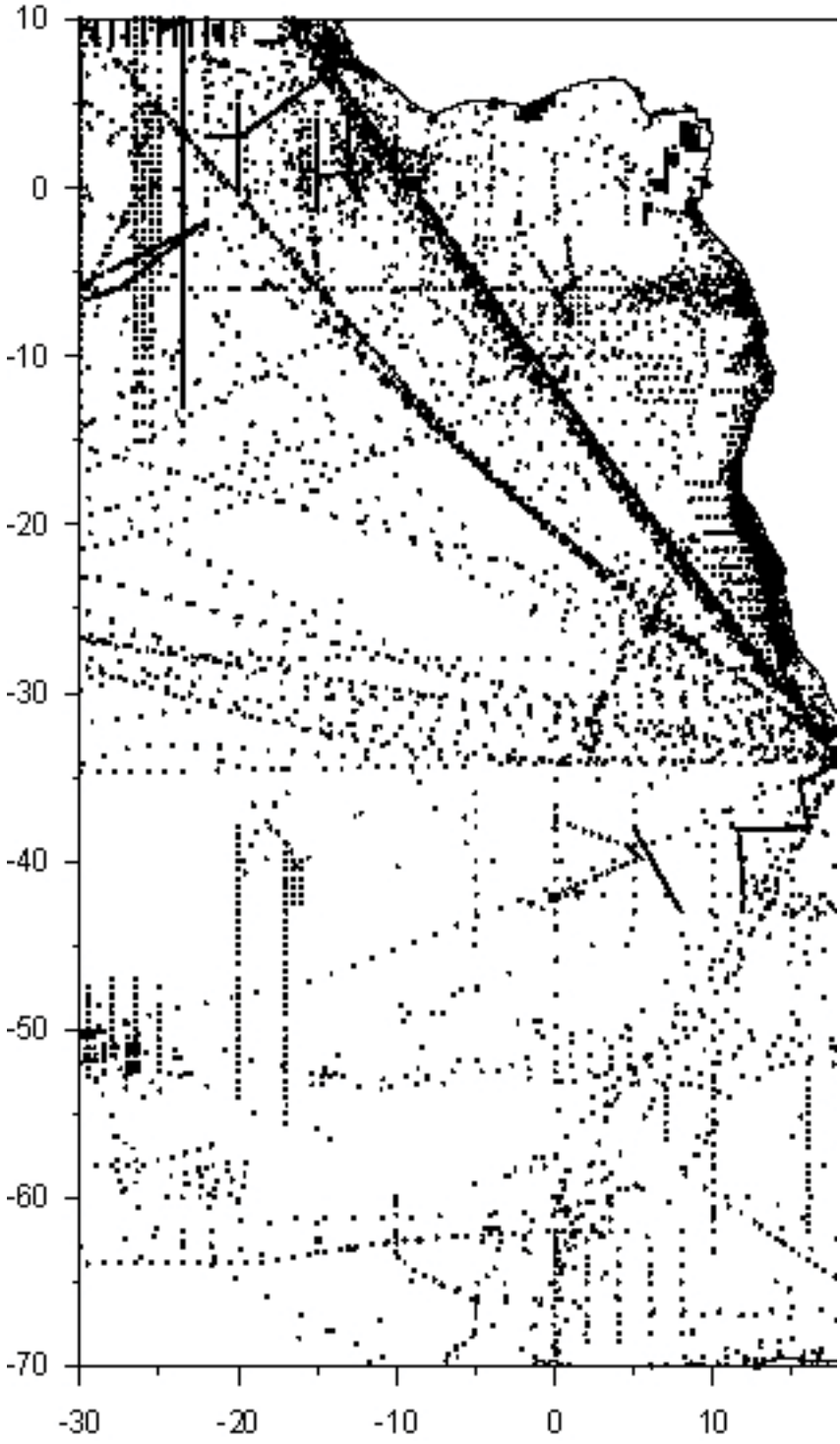
where

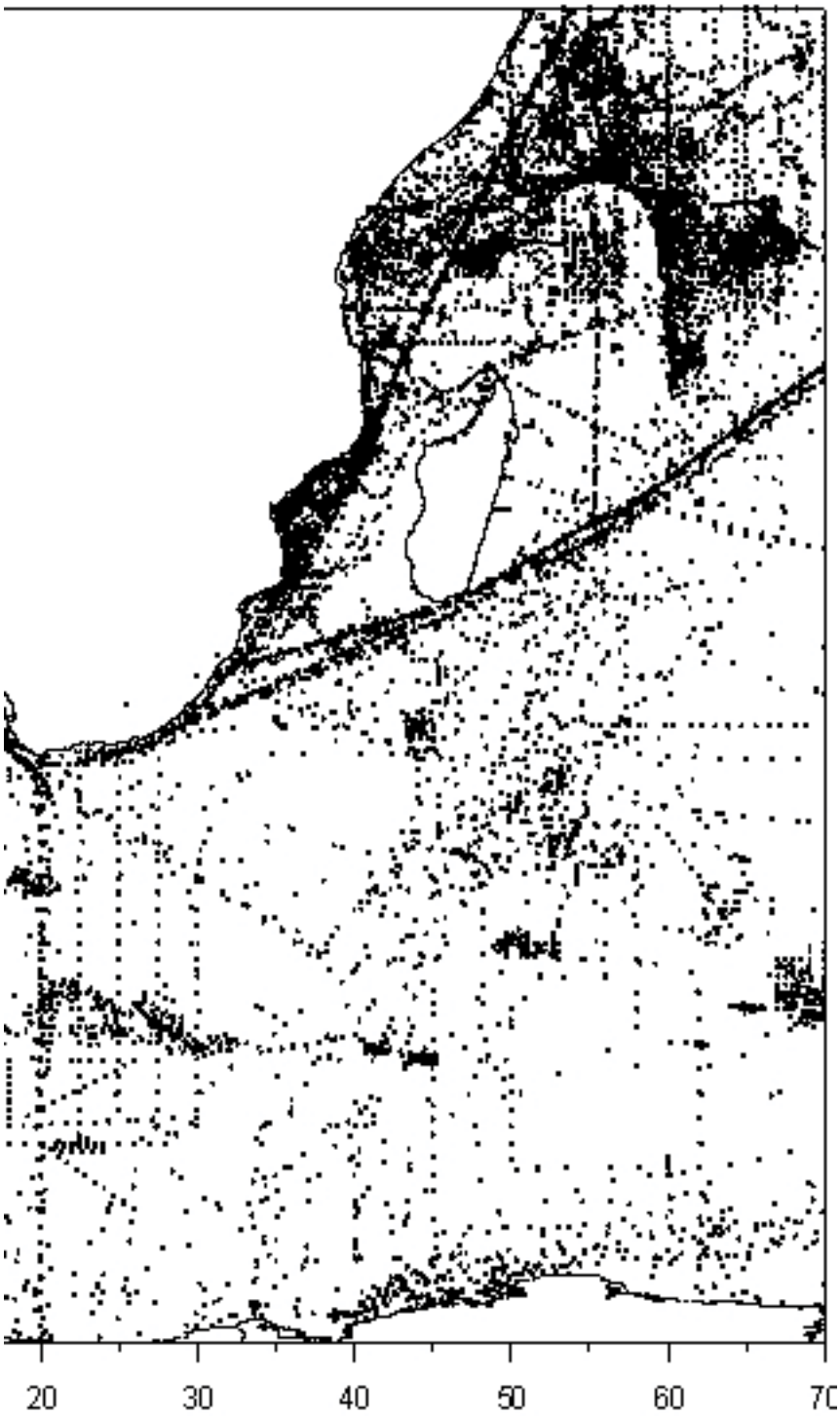
v1 = the value of the variable at the previous depth level
v2 = the value of the variable at the current depth level
v3 = the value of the variable at the next depth level

Table 1:
Test values used for the spike and gradient tests

Variable	$V_{lower\ limit}$	$V_{upper\ limit}$	$V_{threshold}$	V_{grad}
temperature	-10ºC	10ºC	2ºC	10ºC
salinity	-5 PSU	5 PSU	3 PSU	5 PSU

Fig. 2 Plot of hydrographic stations (including CTDs, bottle data, XBTs and MBTs) loaded from WOD2005.




20
30
40
50
60
70

Quality control on vertical hydrographic profiles *(continued..)*

3. Subsurface observation flags

a) Depth inversion and depth duplication

A depth inversion during a down cast occurs when an observation has a shallower depth than the observation directly preceding it. A depth duplicate occurs when a reading has the same depth as the reading immediately before it. In either case the second observation is flagged.

b) Density inversion

Density is calculated using the standard density equation. The density difference between two observations is taken after the deeper observation is adiabatically displaced to the level of the shallower depth. An **inversion** occurs when the difference is less than zero. However, only **excessive** inversions are flagged. To check for excessive inversions, the *gradient* between the two observations is calculated (see (d) below), and excessive inversions are defined as follows:

- for observations < 50m, a gradient > 3×10^{-5} g cm^{-5}
- for observations between 50m and 400 m, a gradient > 2×10^{-5} g cm^{-5}
- for observations > 400m, a gradient of 10^{-6} g cm^{-5}

c) Range

Range checks are used to screen the data for extreme values. WOD has established broad ranges as a function of depth and oceanic basins (shown in Fig. 3) for each variable irrespective of season or year, and an observed value falling outside these ranges is flagged.

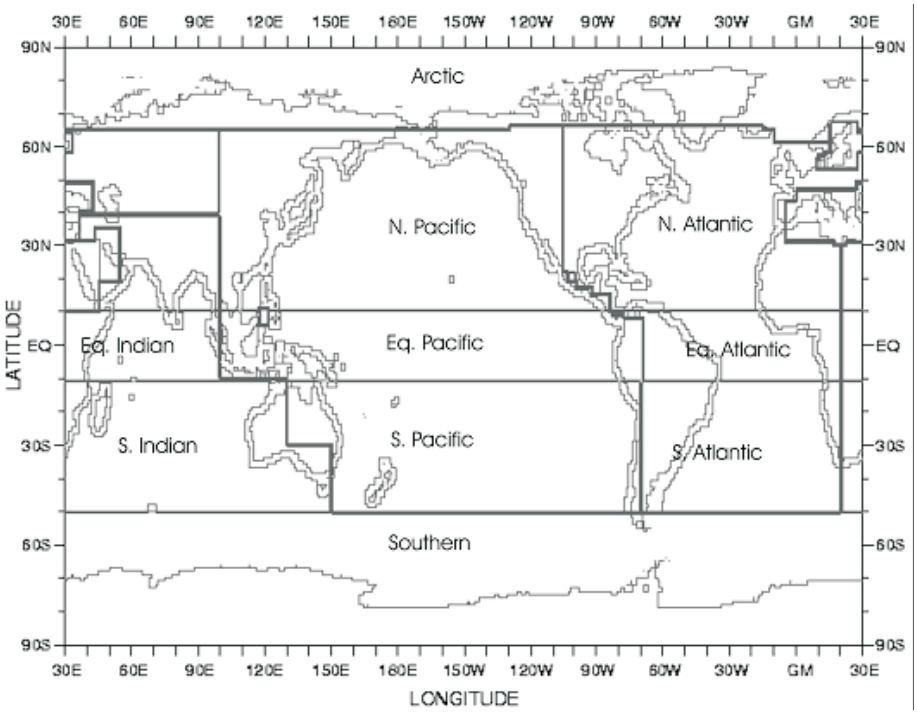


Fig. 3 Geographic boundaries of ocean basin definitions in WOD05.

Quality control on vertical hydrographic profiles *(continued..)*

d) Excessive gradient

For each variable in Table 2 a check is made for "excessive decreases and increases in a value over a depth range". A gradient is defined as:

$$\text{gradient} = (v2 - v1) / (z2 - z1),$$

where
v1 = the value of the variable at the current depth level
v2 = the value of the variable at the next depth level
z1 = the depth (meters) of the current depth level
z2 = the depth (meters) of the next depth level

When dealing with high-resolution instruments a minimum depth difference of 3.0 meters is used when calculating the gradients. If a parameter value exceeds the given limits (see "MGV" Table 2) in a negative sense it is referred to as an "excessive gradient", and if it exceeds the given limits (MIV) in a positive sense it is referred to as an "excessive inversion".

When an excessive **gradient** is located close to an excessive **inversion** (within 8 observations of each other), the observations in between are flagged as having failed the **combined** gradient and inversion check (See Table 3).

Table 2: Maximum gradient and inversion factors

Variable	MGV (Z < 400m)	MGV (Z > 400m)	MIV (Z < 400m)	MIV (Z < 400m)	Unit
Temperature	-0.7	-0.7	+0.3	+0.3	ºC m^{-1}
Salinity	-9.0	-0.05	+9.0	+0.05	PSU m^{-1}
Oxygen	Gradient and inversion checks not made				
Phosphate	-1.0	-0.5	+1.0	+0.5	µM ℓ^{-1}
Silicate	Gradient and inversion checks not made				
Nitrate	-1.0	-0.5	+1.0	+0.5	µM ℓ^{-1}
pH	-0.4	-0.2	+0.4	+0.2	m^{-1}
Chlorophyll	Gradient and inversion checks not made				
DIC	Gradient and inversion checks not made				

Quality control on vertical hydrographic profiles *(continued..)*

Table 3: Comparison of WOD and SADCO Quality Flags.

	WOD Quality Flags	Flags applied by SADCO
Flags for entire profile (as a function of variable)		
0	accepted profile	accepted profile
1	failed annual standard deviation check	failed annual standard deviation check
2	two or more density inversions	two or more density inversions
3	flagged cruise	
4	failed seasonal standard deviation check	
5	failed monthly standard deviation check	failed spike test
6	failed annual and seasonal standard deviation check	
7	bullseye from standard level data or failed annual and monthly standard deviation check	
8	failed seasonal and monthly standard deviation check	
9	failed annual, seasonal and monthly standard deviation check	
Flags on individual observations – depth flags		
0	accepted value	accepted value
1	duplicates or inversions in recorded depth	duplicates or inversions in recorded depth
2	density inversion	density inversion
Flags on individual observations – variable flags		
0	accepted value	accepted value
1	range outlier (outside of broad range check)	range outlier (outside of broad range check)
2	failed inversion check	failed inversion check
3	failed gradient check	failed gradient check
4	observed level “bullseye” flag and zero gradient check	
5	failed combined gradient and inversion checks	failed combined gradient and inversion checks
6	failed range **and** inversion checks	failed range **and** inversion checks
7	failed range **and** gradient checks	failed range **and** gradient checks
8	failed range **and** questionable data checks	
9	failed range **and** combined gradient and inversion checks	failed range **and** combined gradient and inversion checks