GAO: an environmental database and software designed for fisheries biologists

Francis MARSAC[1]

Abstract

This paper presents a software (GAO) processing oceanographic data for fisheries research. GAO is built from large datasets covering the tropical Atlantic and Indian Oceans, some records starting in the early years of the 20th century. Observations are restricted to the upper 500m of the ocean, which encompasses most of the habitat of tunas and other top predators of the offshore pelagic ecosystem. The datasets include oceanographic stations (temperature, salinity, dissolved oxygen, phosphates and nitrates at depth levels), vertical temperature profiles (MBTs, XBTs, CTDs), outputs of the OPA ocean circulation model (temperature, salinity, current vectors, vertical velocity, at 13 depth levels), sea surface temperature and wind-derived fields, Topex sea level anomalies and soon, a temperature-at-depth database at grid points on the global ocean. These datasets are prepared in a way enabling data storage and processing on the local hard disk of a standard personal computer. A Windows® software performs extraction and queries on these databases. The goal of GAO is to be an easy-to-handle tool for working groups in the fisheries domain, where quick reference to physical environment is sometimes required.

Résumé

Ce papier décrit un logiciel (GAO) dédié au traitement de données océanographiques utiles en recherche halieutique. GAO est construit à partir de volumineux jeux de données couvrant les régions tropicales des océans Atlantique et Indien, les premiers enregistrements archivés remontant au début du 20e siècle. Les observations sont limitées à la couche des 500 premiers mètres de l'océan, qui englobe l'essentiel de l'habitat des thons et autres prédateurs océaniques supérieurs. Les jeux de données comprennent des stations océanographiques (température, salinité, oxygène dissous, phosphates et nitrates à différentes immersions), des profils verticaux de température (MBT, XBT, CTD), des sorties du modèle de circulation océanique OPA (température, salinité, vecteurs de courant, vitesse verticale, à 13 niveaux de la colonne d'eau), des champs de température de surface et de vent selon un réseau maillé, les anomalies de niveau de l'océan mesurées par le satellite Topex, et bientôt une base de données de température à des immersions standard selon un réseau maillé sur l'océan mondial. Ces jeux de données sont préparés de manière à permettre l'archivage des données et leur traitement en mode local, sur le disque dur d'un ordinateur de bureau. Un logiciel développé sous Windows© effectue des extractions et des requêtes sur ces bases de données. Le but de GAO est de devenir un outil d'emploi aisé dans les groupes de travail sur les pêcheries où des références rapides à l'environnement physique sont souvent demandées.

[1] IRD, UR 109 Thetis, Centre de Recherche Halieutique Méditerranéen et Tropical, BP 171, 34203 Sète cedex, France. E-mail : marsac@ird.fr

Introduction

GAO is the French acronym for "Gestionnaire d'Applications Océanographiques", namely a package for processing ocean data. GAO comprises several environmental data sets and a user-friendly interface to undertake data extraction and processing from the data sets. It has been designed for fisheries biologists wishing to integrate environmental indices, ocean and atmospheric data into stock and population analyses. To be fully operational and convenient for most of the users, GAO is installed on PCs, the data sets being compressed and organized in a way enabling a quick execution of the queries.

There are plentiful oceanographic observations in the world ocean. For years, they have been archived and managed with little coordination among the different national data centers. During the last decade, a tremendous effort of data compilation has been made by the major research institutions to make this information available on the web (see Fromentin et al 2005, this meeting, for a selection of the major sites). However, this new way of getting data does not provide an optimal solution to the user. Most often, the sites are dedicated to a given type of data (*in situ* observations, remote sensing data and model outputs). Therefore it can be time consuming to build a comprehensive set of observations in a restricted area. Moreover, the parameters can be available in different units among web sites (especially for nutrients). Finally, because of a variety of formats used, it is often necessary to manipulate data prior to have them usable for a specific analysis.

Several products have been released on CD-ROMs in the recent years to overcome some of these constraints. One of the most well-known is the COADS (see Slutz *et al* 1985, Woodruff *et al* 1987 for data; Mendelsohn and Roy 1996 for the CODE software). COADS is restricted to surface parameters, i.e. gridded fields (2° x 2°) of sea surface temperature (SST), wind and wind-derived parameters. Another product, the CLIMLAB2000 of the of International Research Institute for Climate Prediction (IRI 1999), is a powerful statistical package dedicated to ocean data analysis, but it only holds SST fields and time series of atmospheric indices.

GAO is also intended to make life easier to biologists who are not too familiar with ocean data processing. It is not duplicating the other products as GAO mainly deals with subsurface data, the water column from surface to 500 m depth, and manages discrete observations, gridded fields and model outputs. Two major oceanic areas are covered: Atlantic Ocean, from 50°N to 30°S, and Indian Ocean, from 30°N to 60°S.

Data collection and preparation

Different data sources were used to prepare the GAO datasets. The objective of gathering data from the beginning of the 20th century was targeted. Six segments of data are presently managed by GAO:

a) Discrete observations:

- Oceanographic stations
- Vertical temperature profiles

b) Gridded fields

- Wind stress
- Sea Surface temperature
- OPA circulation model
- TOPEX/POSEIDON sea level anomalies

The main characteristics of the databases are summarized in Table 1. The major tasks undertaken in the data preparation are quality control/validation (for discrete observations) and restructuring of datasets to speed up the execution of the queries.

1. Oceanographic stations

1.1. Sources

Data were initially obtained from the US National Oceanographic Data Centre (NODC) and the French *Banque Nationale de Données Océanographiques* (BNDO). This was undertaken in 1988 when the compilation of the GAO datasets was initiated. In 1995, after the release of the World Ocean Atlas 1994 of Levitus and Boyer (1994) (set of 9 CDs), the whole database was updated to include the recent years. At the same time, old data that were not present in the former data sets were included. The data series covers the period 1906-1991 for Atlantic and 1906-1994 for Indian Ocean. The update until 2001 based on the latest release of the World Ocean Atlas 2001 (Stephens et al 2002, Boyer et al 2002, Locarnini et al 2002, Conkright et al 2002) is underway.

1.2. Quality control

The parameters included in the datasets are temperature (in °C), salinity (in $g.l^{-1}$), dissolved oxygen (in $ml.l^{-1}$), phosphates and nitrates contents (in μM) from 0 to 500 m. The data available on the World Ocean Atlas have been screened for validation. Observations are flagged to indicate the degree of confidence about the climatology at different time scales (month, season, year). The quality control procedures applied to data of World Ocean Atlas are fully described on the NODC web site (http://www.nodc.noaa.gov/OC5/WOD01/doc01.html). Briefly, the sequence of controls is performed to check distance between stations (through a speed check), duplicate records (identical profiles, identical stations and overlapping cruises), depth inversion, range of values, excessive gradients…). The flags assigned by the NODC were kept in the final GAO dataset.

1.3. Dataset configuration

The previous tasks are performed on ASCII files. However, it was necessary to reorganize the datasets in order to enable an immediate and direct access to the data and to minimize the size of the database. Then, we designed tables of pointers indexing the database. Pointers make the link between the two main segments of the database: headers (metadata) and observations at depth. The second segment of the database is converted into unformatted binary form (as well as some intermediate tables) so that information is very compact and can be managed on a standard PC.

The datasets resulting from this processing cannot be modified by the user. New data cannot be appended either by the user. The datasets are updated once a year by the data manager and distributed to the users.

1.4. Query protocol

A query consists in the selection of an area, a period, one or more parameters and a range of quality flags for these parameters). When a query is submitted by a user, the sequence of tasks is the following :

a) the table containing year-month pointers is accessed first: the result is a selection of headers (H1) matching the time frame;

b) the headers H1 are read to select those matching the area, the parameters and the quality flags: header list H2.

c) the pointers of the headers H2 link to record numbers of an intermediate table. Each record of this intermediate file corresponds to a single station/profile; it contains the records to read in the segment "observations at depth".

The observations at depth are read and returned to the processing job.

1.5. Data coverage

In the Atlantic Ocean, a total of 78 724 oceanographic stations have been archived in GAO in the area 80°W-20°E / 30°N-30°S for 1906-1991. Table 2 shows the number of stations by year and month; Figure 1 is a series of quarterly maps of the stations. The area is presently inter-tropical; soon, it will be enlarged to temperate waters, to the north and the east (up to latitude 50°N and longitude 0°).

In the Indian Ocean, there are 13 191 stations in the Western basin (30°E-80°E / 30°N-30°S) for 1906-1994. Table 3 shows the number of stations by year and month; Figure 2 is a map of the total distribution of stations. An extension to the whole ocean (longitude 120°E and latitude 60°S) is underway.

2. Vertical temperature profiles

2.1 Sources

The first dataset processed in 1988 was provided by the TOGA Subsurface Data Centre (TSDC) settled in Brest, France. This centre is now the SISMER, still on the IFREMER site in Brest. The bulk of the data is composed of XBT casts which were deployed from voluntary operating ships since 1985, in the framework of TOGA. The data density is very satisfactory along shipping lines, this configuration being appropriate for time series analyses in these sub-regions. Other profiles are from CTD casts made by research vessels of various research institutions. We considered the observed profiles (instead of interpolated temperature at standard depths), which makes a variable and numerous values on the water column (an XBT logs data approx. every 3 meters).

In order to have a comprehensive temperature-at-depth (TAD) database, we appended the observations of the oceanographic stations to the bulk of XBT and CTD datasets. The old data are dominated by mechanical bathy thermographs (MBT) or Nansen bottles. The number of observations-at-depth is not as large as that recorded by XBTs or CTDs.

The data obtained from SISMER and other sources are in ASCII; a processing is performed to convert these data into the GAO format.

2.2 Quality control

The data consists of only two parameters, depth and temperature. The first series of controls checks positions and duplicate profiles. Then validation of the TADs is carried out using confidence intervals about the Levitus monthly-based climatology at standard depths. Quality flags (1, 3 or 4) are assigned to each TAD, by comparing the observed value with a standard deviation multiplier which defines a confidence interval :

- for flag 1 (good quality): a temperature must be within 2.5 standard deviation in the layer 0-125 m and within 3 standard deviation in 125-500 m;
- if the temperature is off the bounds, we check the flag 3 (doubtful): the value must be within 3 standard deviation in the layer 0-125 m and within 3.5 standard deviation for 125-500 m;
- the flag 4 (bad) is assigned when none of these checks are conclusive.

The overall quality flag of the profile depends on the distribution of individual flags at depth. A quality flag of 1 requires 100% of TAD flagged 1. If some flag 3 and less than 10% of flag 4 are present in the individual measurements, the overall flag will be 3. If there are

more than 10% of flag 4, the profile will be flagged 4. On the overall database, there are more than 80% of profiles with flag 1.

2.3 Dataset configuration

A protocol similar to that of the oceanographic stations is applied to the TAD dataset: pointer tables are generated and data are converted into unformatted binary. Consequently, the database cannot be modified or appended by a user.

2.4 Query protocol

The TAD database having the same structure as the Oceanographic stations, the way data are accessed by a query is similar to that described in 1.4.

2.5 Data coverage

In the Atlantic Ocean, 61 865 profiles are stored for the sole period 1985-2000. Integrating the historical data set (prior to 1985) and updating data up to 2003 is underway. The area presently covered is 80°W-20°E / 30°N-30°S, but a extension of the northern boundary up to 50°N and longitude 0° is also underway. This will result in a tremendous increase of the database as more than 22 000 profiles have been collected in the area 30°N-50°N. Table 4 shows the current number of stations by year and month; Figure 3 presents maps of the quarterly distribution of profiles.

In the Indian Ocean, data cover the period 1906-2000: the number of temperature profiles is 107 154 in the area 20°E-120°E / 30°N-60°S. Table 5 shows the number of stations by year and month; Figure 4 is a map showing the quarterly distribution of the profiles.

3. Pseudo wind stress

3.1 Sources

Data used in GAO are downloaded from the site of the FSU, Florida State University (http://www.coaps.fsu.edu). Data for Atlantic are processed by IRD (Servain et al 1987, 1997) and are also available on the FSU web site. Data for the Indian Ocean are processed by the FSU (Legler et al 1989, 1997). The original data analyzed by both institutions are the *in situ* observations (merchant ships, buoys, and other marine observing stations) that number about 165 000 for any given month for the entire work. The data files downloaded and converted into GAO format are gridded fields by month. The grid size is 2° for the Atlantic and 1° in the Indian Ocean.

3.2 Quality control

The quality control and data analysis performed on original observations is well described on the FSU web site. Here follow the guidelines provided by the authors. Individual reports of wind speed are checked: if the speed exceeds 40 m s^{-1}, the report is deleted. The individual wind vector data are first converted to pseudo-stress. A pseudo-stress vector is defined as the wind components multiplied by the wind magnitude. A pseudo-stress vector of 60 m^2s^{-2} corresponds to 1 dyne cm^{-2} assuming a drag coefficient of 1.4 x 10-3 and an air density of 1.2 kg m^{-3}. The pseudo stress reflects the force exerted by the wind on the upper layer of the ocean; it can provide an estimate of turbulence, an environmental proxy used in marine ecology.

Objective filtering is then applied to the data to eliminate questionable pseudo-stress reports. For each month, x and y components of each report are checked within boundaries about the basin-wide average. Values falling outside +/-30 times the average, the observations are deleted. Less than 1% of all reports are removed.

The remaining observations are then averaged in regular geographic boxes. The final screening deals with each box individually. In boxes with greater than 5 observations per month, individual observations are removed when either component is greater than +/-3.9

standard deviations from the average value. Again, typically less than 1% of the reports are removed. The remaining reports are averaged onto a 2° grid for Atlantic and 1° grid for Indian, with empty boxes primarily in areas outside of the shipping lanes. Finally a variational direct-minimization objective analysis is performed on these data to provide the resulting fields.

3.3 Dataset configuration

The data are organized in annual files. As the size of each file is not big (c.a. 130 Kb for Atlantic and 700 Kb for Indian), it was not necessary to create pointer tables. The records managed by GAO are binary unformatted and in direct access. A pseudo wind stress climatology in the same grids (2° and 1°) is also included in the database to compute anomalies from the queries.

3.4 Query protocol

The protocol is simpler than in the two previous datasets. The files to be read are only those of the years selected by the user. Once a file opened, the geographical box is read in each record and the wind stress values are returned and processed if the box is included in the area selected by the user.

3.5 Data coverage

In the Atlantic, the 2° grid field of pseudo wind stress covers a large portion of the tropical ocean: 60°W-16°E / 30°N-20°S, for the period 1964-2003.

In the Indian Ocean, data cover the whole tropical region: 30°E-120°E / 24°N-30°S, for the period 1970-2003.

4. Gridded Sea Surface temperature

We shall not detail this dataset, as it is strictly similar to the pseudo wind stress one. The sources are the same (ship and buoys reports). The data set comprises a single parameter (SST) instead of two (x and y component of the wind stress) but the dataset exhibits the same configuration.

5. OPA circulation model

5.1 Source

OPA (an acronym for "Océan PArallélisé"), an Ocean General Circulation Model (OGCM) developed at the Laboratoire d'Océanographie DYnamique et de Climatologie (LODYC), University of Paris-6, is intended to be a flexible tool for studying the ocean and its interactions with the others components of the earth climate system (atmosphere, sea-ice, chemical tracers, ...) over a wide range of space and time. A complete description of the model is given by Madec et al (1998). The document can be downloaded on the site http://www.lodyc.jussieu.fr/opa/. This model provides a good opportunity to study the time-space variability of some parameters related to the ecology of large migratory species. This product is now hosted by the Mercator project (http://www.mercator-ocean.fr) and numeric outputs can be downloaded by registered users.

5.2 Dataset configuration

We intended to make available into GAO a limited number of parameters (6: depth-temperature-salinity, x and y components of the current, vertical velocity) distributed on 13 levels (5, 15, 25, 36, 51, 70, 90, 110, 130, 150, 172, 217, 325 m) on a grid size of 1° - fortnight. The initial resolution of data gathered being 0.5° lat-lon and 5 days, we have pooled the data on a 1° lat-lon fortnight basis using bin averages. We kept the same configuration as that used for oceanographic stations and vertical temperature profiles. There are headers describing the location (1° box) and date of the stratum, table of

pointers to speed up the search of the records matching the query in the observations-at-depth database. This database is in direct access and binary unformatted.

5.3 Query protocol

The protocol is similar to that described in paragr. 1.4.

5.4 Data coverage

Only the Atlantic Ocean is concerned with OPA model in the GAO database. The area covered stretches over the tropical region, 80°W-16°E / 20°N-20°S for the period 1980-2000. An update up to 2003 is planned during the first semester of 2005. Although aggregating data to 1°-fortnight and restructuring the dataset, the OPA model output remains a huge database (530 Mb) which is read by GAO from a CD-Ro

6. Topex-Poseidon sea-level anomalies

6.1 Sources

The sea level data base used is downloaded by ftp (ftp.csr.utexas.edu/pub/sst/tp_data) from the site of the University of Texas at Austin Center for Space Research (UT/CSR). This Centre offers a free access to data. This database contains gridded sea level anomaly heights observed by the TOPEX/POSEIDON (T/P, 1992-Present) satellite altimeter. The sea level anomaly (SLA) is the deviation of the sea surface about a mean surface. Data are stored on a 1° grid according to the T/P repeat cycle number (10 days). The raw 1-second (7 km) data are used to compute the smoothed 1° grid directly. A basin check is also made to ensure that, for instance, data from the Gulf of Mexico is not used in the computation of the grid in the eastern Pacific, or vice versa. A smoothing radius of 500 km is used. Weighted averages are computed for each grid, based all points within the basin and 500 km radius, where the weight = 4.0*exp(-(radius/500)**2. The estimated accuracy of the sea level anomalies in this database is 3 to 4 cm over the entire ocean.

6.2 Dataset configuration

We applied the configuration using pointer tables, direct access to the core database and conversion of values in unformatted binary. The header file contains metadata for each cycle (every 10 days) and pointers making a link to the SLA dataset.

6.3 Query protocol

A query is based on an area selection and on a time window frame, which is defined according to the 10-days cycle. For a given month, the user is prompted to choose the cycle of interest (in a pop up list) and the relevant data for this cycle are read and returned to the processing job. Headers are ASCII files whilst the SLA dataset is in direct access and unformatted binary.

6.4 Data coverage

This dataset covers the global ocean, for the period Sep.1992 – Oct. 2001. An update to 2003 is planned for the first semester of 2005. The actual size of the restructured database within GAO is 235 Mb, which is not that excessive when we consider the number of strata to manage (1° boxes by 10-days over 10 years).

System organization

GAO is installed in a directory under the root of a local disk of a PC. The main directory contains the executable and information files handled by the system. Below are 2 sub-directories: DATA and USERS.

a) The DATA directory holds the data sets, subdivided into 6 components (each of them being a directory):

- CARTO contains the coastline files
- HYDRO contains the oceanographic data sets
- FSU contains the Indian Ocean Wind stress fields
- VTS contains the Atlantic Ocean Wind Stress fields
- SST contains the Atlantic Ocean gridded sea surface temperature
- TOPEX contains the sea level anomalies measured by Topex-Poseidon over the global ocean

b) The USERS directory holds sub-directories where the output files resulting from the queries are stored. When opening a GAO session, a user ID is requested; this ID will be the name of the default directory where the files of the running session will be written. The system indicates whether the ID sub-directory exists or not. If not, the ID sub-directory is created. This allows the user to organize and keep track of its work over time and prevents from having output files created from various sessions mixed in a single output directory.

Software

The GAO software has been developed in Visual Basic to use the Windows environment. The main parameters of a query (selection of area and time windows, quality flags, parameters…) are managed by Visual Basic modules. Then, the queries are executed in the background by routines written in FORTRAN.

In the present version (4.0), the outputs of the queries are essentially tabulated text files (ASCII) which are written in the default output directory (user ID). They can easily be imported in spreadsheets, graphics or statistical packages for further processing. A general flowchart of the various procedures is given in Figure 3.

After entering the user ID, the main GAO window appears on the screen. Three data categories are displayed as pop off menus: Stations, temperature profiles, gridded fields. They refer to the datasets detailed in the previous sections. In the bottom of the window, there are three icons: "Edit" to launch an editor (Professional Editor, a freeware), "View log file" to trace the sequence of operations undertaken under the user ID session, and "Reset log file" to erase the sequence.

The queries on a given type of data are launched using the pop-off menus. They are briefly described hereafter.

- *Stations*

Settings: it is the first item of the pop-off menu, where the user is guided to select an ocean basin (Atlantic or Indian), the selection of a time window (year - month) and an area (latitude and longitude boundaries). This makes the default selection for all queries until any modification of the selection.

Data summary : produces a table showing number of stations by year and month, either in the whole ocean basin or in the area selected by the user (cf. Settings)

Data table: produces a listing of values at depth (temp, salinity, DO, PO4 and NO3) for all the stations matching the user's request.

Mapping: produces a map of the area with location of stations at different time steps (from fortnight to year). It is possible to zoom in or out, and plot the vertical profile of all available parameters for a selected station.

Statistics : this item opens to 3 sub-levels :

- By station: produces a table of values at different depth levels (0, 100, 200, 300, 400 m) for each station matching the space-time selection.
- Quick Plot: plots on a map the location of the stations written during the query "By Station"
- Calculate : this item opens to 2 sub-levels
    - Average by 1° square: computes the average for each parameters at the 5 depth levels (0 to 400 m) across a 1° grid, on the file created by the query "By Station"
    - Time series: computes a time series (average over a user-defined area) of a parameter at a given depth, from the file created by the query "By Station". The mean value, standard deviation an number of observations at each time interval are written. The time interval of the series can be either fortnight or month. It has to be consistent with that of the source file.

❑ *Temperature profiles*

Settings: same as above

Data summary: same as above, applied to temperature-at-depth data

Mapping: same as above

Statistics: this item opens to 3 sub-levels:

- By profile : produces a table of SST, temperature at 100 m, depth of mixed layer, value and depth of the peak temperature gradient, and the depth of isotherms (from 30°C to 10°C) for each profile matching the space-time selection.
- Quick Plot : plots on a map the location of the profiles written during the query "By profile"
- Calculate : this item opens to 4 sub-levels
    - XYZ table: extracts the values of a given parameter (among those produced by the query "By profile") and writes them together with their position (useful for contours or spatial analysis).
    - Average by 1° square: computes the average for each parameter extracted by the query "By profile" across a 1°-grid.
    - Time series: computes a time series (average over a user-defined area) of a parameter at a given depth, from the file created by the query "By Station". The mean value, standard deviation an number of observations at each time interval are written. The time step of the series can be either fortnight or month. It has to be consistent with that of the source file.
    - Space-time series: computes a table of average values across a time step (month or fortnight) and a space variable, either longitude or latitude. All the parameters extracted by the query "By profile" are considered. The space interval is by 1° (useful for plotting Hovmoller diagrams).

❑ *Gridded fields*

OPA model

- Settings: same as above. It is necessary to specify the letter of the CD drive as OPA data are read from a CD-Rom (it is normally set automatically but still good to check).

- X-Y-Z table: extracts the values of a given parameter (out of 5) at a given depth, by fortnight or by month. Values are written with their date and location (useful for contours or spatial analysis). All the oceanic strata are documented (no blanks).
- Parameters and anomalies: computes an average and standard deviation across a 1°-grid for a parameter for a selected time window. Values are written with their location. It is also possible to compute anomalies : the years used for calculating a climatology have to be specified: the table written holds the climatological value, the parameter value and its anomaly with location of the 1° boxes.
- Vertical profiles: computes an average profile (with standard deviation) for all parameters at each depth level in the area selected.

Atlantic SST

- Settings : same as above
- X-Y-Z table : extracts all SST and SST anomalies (SSTA) across a 2°-grid (Atlantic) or 1°-grid (Indian) by month, matching the space-time selection
- Anomaly indices: computes the differences of SST anomalies between two areas. The result is a time series where SST and SSTA of both area are displayed, along with the anomaly differences which makes a dipole index.

Pseudo Wind Stress

- Settings: same as above.
- X-Y-Z table: extracts all X and Y components of the pseudo wind stress (PWS) vectors across the grid and by month. The wind speed (ms^{-1})and the turbulence (m^3s^{-3}) are appended to their corresponding X and Y PWS at each grid point.
- Parameters and anomalies: averages values of parameters (wind stress, scalar wind, turbulence) or their anomalies across the grid. Results include standard deviation. They are written along with the location of each grid point.

Topex sea level anomalies

- Settings: same as above
- File for a map: produces a table of sea level anomalies (SLA) at each 1°-grid point for a given TOPEX cycle selected by the user.
- Time series: produces a time series of SLA in a selected area. SLA at grid points are averaged over the entire area for each 10-days interval.

Temperature at standard levels

The set of queries corresponding to this database are under development.

Perspective

GAO is a tool written by researchers for research. This implies that new modules can be appended to the current set of queries, or new datasets to be included in the system. We wish to gather feed-back from users to maintain a product that can be helpful notably to fisheries scientists working on high-seas migratory species.

Some improvements are already planned for the next 6 months:

1) in the present version, GAO only writes files to be imported in other packages for plots, maps or statistical analysis. In the next version, there will be plot windows to produce graphs from the data tables produced by the queries. The graphs will be

basic, as lots of sophisticated graphic packages already exist. However, an immediate graph of the data can be very informative.

2) As mentioned in the document, databases need to be updated.

3) the oceanographic stations and temperature profiles databases are made of individual observations at observed levels. The same type of data will be available in gridded fields at standard depths.

4) Time series of large-scale atmospheric indices (difference in Sea-level pressure anomalies, like SOI, IOI or NAO…), and other indices (LOD,…) will be included in GAO.

GAO software and databases should be put on a web site for download. This will facilitate dissemination of updates and return of bug reports. If this product proves to be successful among the users, it could be used during the working groups of the tuna commissions of the Atlantic (ICCAT) and Indian Ocean (IOTC).

References

Boyer, T. P., Stephens, C., J. I. Antonov, M. E. Conkright, R. A. Locarnini, T. D. O'Brien, H. E. Garcia (2002). World Ocean Atlas 2001, Volume 2: Salinity. S. Levitus, Ed., NOAA, Atlas NESDIS 50, U.S. Government Printing Office, Wash., D.C., 165 pp., CD-ROMs.

Conkright, M. E., H. E. Garcia, T. D. O'Brien, R. A. Locarnini, T. P. Boyer, C. Stephens J. I. Antonov (2002). World Ocean Atlas 2001, Volume 4: Nutrients. S. Levitus, Ed., NOAA, Atlas NESDIS 52, U.S. Government Printing Office, Wash., D.C., 392 pp., CD-ROMs.

Fromentin, J-M, F. Royer and F. Marsac (2004). Open environmental databases for open-sea fisheries biologists. ICCAT. SCRS/2004/171, 5 p.

Legler, D. M., I. M. Navon, and J. J. O'Brien (1989).: Objective Analysis of Pseudostress Over the Indian Ocean Using a Direct-Minimization Approach. Monthly Weather Review, 117: 709-720.

Legler, D. M., J. N. Stricherz, and J. J. O'Brien (1997). TOGA Pseudo-stress Atlas 1985-1994, Volume III: Indian Ocean, Florida State University, Tallahassee, FL, 155 pp

Levitus, D. and T.P. Boyer (1994). World Ocean Atlas 1994. US Department of Commerce, NOAA, National Environmental Satellite, Data, and Information Service.

Locarnini, R. A., T. D. O'Brien, H. E. Garcia, J. I. Antonov, T. P. Boyer, M. E. Conkright, C. Stephens (2002). World Ocean Atlas 2001, Volume 3: Oxygen. S. Levitus, Ed., NOAA, Atlas NESDIS 51, U.S. Government Printing Office, Wash., D.C., 286 pp., CD-ROMs.

Madec, G., P. Delecluse, M. Imbard, and C. Lévy (1998). OPA 8.1 Ocean General Circulation Model reference manual. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 11 , 91pp.

Mendelssohn, R. and C. Roy (1996). Comprehensive Ocean Data Extraction Users Guide. U.S. Dep. Comm., NOAA Tech. Memo. NOAA-TM-NMFS-SWFSC-228, La Jolla, CA. : 67 p.

Servain, J., J. N. Stricherz, and D. M. Legler (1997). TOGA Pseudo-stress Atlas 1985-1994, Volume I: Atlantic Ocean, Florida State University, Tallahassee, FL, 158 pp.

Servain, J., Seva, M., Lukas, S. and Rougier, G. (1987). Climatic Atlas of the Tropical Atlantic Wind Stress and Sea Surface Temperature: 1980-1984. Ocean Air Inter. J., 1, 109-182

Slutz, R.J., S.J. Lubker, J.D. Hiscox, S.D. Woodruff, R.L. Jenne, D.H. Joseph, P.M. Steurer and J.D. Elms (1985). Comprehensive Ocean Atmosphere DataSet ; release 1. NOAA Environmental Research Laboratories, Climate Research Program, Boulder,CO : 268 p.

Stephens, C., J. I. Antonov, T. P. Boyer, M. E. Conkright, R. A. Locarnini, T. D. O'Brien, H. E. Garcia (2002). World Ocean Atlas 2001, Volume 1: Temperature. S. Levitus, Ed., NOAA, Atlas NESDIS 49, U.S. Government Printing Office, Wash., D.C., 167 pp., CD-ROMs.

Woodruff, S.D., R.J. Slutz, R.L. Jenne and P.M. Steurer (1987). A Comprehensive Ocean-atmosphere Data set. *Bull. Amer. Meteor. Soc.*, 68 : 1239-1250

Table 1 – Overview of the datasets managed by GAO

Database	Period	No of stations/ profiles	Grid size	Area	Size (Mb)
Atlantic					
Oceanographic stations	1906-1994	78 724	Exact location	80°W-20°E 30°N-30°S	36
Vertical temperature profiles	1985-2000	61 865	Exact location	80°W-20°E 30°N-30°S	76
Pseudo wind stress	1964-2003		2° - month	60°W-16°E 30°N-20°S	5
Sea surface temperature	1964-2003		2° - month	60°W-16°E 30°N-20°S	4
OPA circulation model	1980-2000		1° - fortnight	60°W-16°E 20°N-20°S	530
Indian Ocean					
Oceanographic stations	1906-1994	13 191	Exact location	30°E-80°E 30°N-30°S	6
Vertical temperature profiles	1906-2000	107 154	Exact location	20°E-120°E 30°N-60°S	96
Pseudo wind stress	1970-2003		1° - month	30°E-120°E 24°N-30°S	22
Worlwide					
TOPEX sea-level anomalies	1992-2001		1° - 10 days	global	235
Temperature at standard levels(*)	1955-2003		5°lon x 2°lat and month	global	301

(*) under development, source JEDAC.

Table 2 – Year-month distribution of the oceanographic stations in the Atlantic Ocean

80°W - 20°E / 30°N - 30°S

YR	J	F	M	A	M	J	J	A	S	O	N	D	TOTAL
1906	0	6	8	0	0	0	0	0	0	0	0	0	14
1910	0	0	0	0	9	1	0	0	0	0	0	0	10
1911	0	0	0	0	0	21	15	12	8	4	0	0	60
1913	0	0	0	0	0	5	11	0	1	2	4	10	33
1914	0	9	12	2	0	0	0	0	0	0	0	0	23
1919	0	0	0	0	0	0	0	0	0	0	6	0	6
1921	0	0	0	2	4	0	0	0	0	3	19	1	29
1922	6	14	6	27	2	2	0	0	0	0	0	0	57
1925	7	0	0	0	19	0	3	26	0	0	0	0	55
1926	0	0	0	0	30	7	28	15	68	43	49	50	290
1927	23	44	76	27	10	0	1	9	0	0	0	0	190
1928	0	0	0	0	0	0	0	13	9	4	0	0	26
1930	0	2	7	0	0	0	3	0	0	0	0	0	12
1931	0	0	0	3	20	0	0	0	0	12	0	0	35
1932	7	0	47	10	0	0	0	0	0	0	0	0	64
1933	0	13	120	55	5	0	0	0	0	0	0	8	201
1934	3	34	0	3	2	0	2	0	0	0	7	1	52
1935	0	0	0	1	0	0	0	0	0	7	4	0	12
1936	3	1	1	5	0	0	0	0	0	0	0	0	10
1937	81	0	0	63	2	0	0	0	0	0	0	0	146
1938	0	68	21	10	7	0	0	0	0	0	0	0	106
1939	46	88	72	11	15	0	0	0	0	0	0	0	232
1940	0	73	0	0	0	0	0	0	0	0	0	0	73
1941	10	38	0	0	0	0	0	0	0	0	0	0	48
1942	0	0	1	0	0	0	0	0	0	0	0	0	1
1945	0	0	0	0	0	0	0	1	0	0	0	0	1
1946	0	4	0	0	0	0	0	0	0	30	17	0	51
1947	10	0	1	31	0	16	0	3	18	0	0	0	79
1948	0	0	0	0	0	5	9	10	9	147	237	41	458
1949	9	8	23	21	20	2	19	0	0	0	1	21	124
1950	0	53	49	28	4	0	0	0	27	17	39	62	279
1951	0	0	64	21	0	0	0	0	4	0	0	0	89
1952	14	93	101	60	10	3	0	13	19	32	58	19	422
1953	7	19	45	66	20	33	38	2	4	45	13	0	292
1954	20	18	3	27	0	62	4	0	10	1	45	44	234
1955	4	18	14	3	4	31	2	6	7	4	17	9	119
1956	13	18	23	2	32	41	24	12	9	37	61	19	291
1957	8	39	78	61	54	12	20	32	28	140	68	14	554
1958	34	98	32	26	47	46	36	118	68	140	242	42	929
1959	57	23	58	63	47	46	51	43	11	153	86	36	674
1960	73	119	195	51	106	18	58	91	34	82	62	17	906
1961	75	38	170	142	53	70	50	80	146	59	263	146	1292
1962	11	46	388	73	146	104	31	64	36	167	87	18	1171
1963	38	231	473	338	94	31	24	331	281	110	146	43	2140
1964	56	207	190	170	92	52	149	178	258	171	164	180	1867
1965	214	251	223	125	264	175	123	173	265	164	203	88	2268
1966	125	274	377	229	150	219	70	238	221	243	166	110	2422
1967	81	137	159	183	137	80	38	183	120	277	221	117	1733
1968	129	321	281	337	399	227	115	114	104	416	353	69	2865
1969	233	280	172	291	905	738	822	163	151	322	188	252	4517
1970	168	357	245	306	261	187	184	226	347	293	293	149	3016
1971	49	159	165	340	265	302	158	179	205	294	398	215	2729
1972	246	267	470	438	90	230	786	823	249	224	191	362	4376
1973	476	529	453	263	410	424	200	364	169	221	320	55	3884
1974	180	264	226	125	211	877	3010	3267	3035	70	63	25	11353
1975	129	208	89	45	80	77	53	49	67	108	125	83	1113
1976	55	165	209	94	39	37	67	26	71	136	158	41	1098
1977	53	26	83	51	118	237	167	232	98	221	147	155	1588
1978	83	35	91	184	64	122	168	216	58	214	206	105	1546
1979	371	368	451	157	545	439	299	264	47	209	160	174	3484
1980	131	210	272	124	277	111	170	138	81	114	151	51	1830
1981	87	79	115	99	120	81	176	180	100	171	402	130	1740
1982	111	85	87	263	25	40	217	134	68	121	100	76	1327
1983	201	237	267	132	378	82	241	254	45	156	103	17	2113
1984	127	259	97	86	76	10	96	92	175	28	183	144	1373

Table 2 (continued)

YR	J	F	M	A	M	J	J	A	S	O	N	D	TOTAL
1985	93	256	161	182	165	261	113	220	108	45	416	25	2045
1986	234	231	139	315	168	71	75	161	84	89	237	211	2015
1987	17	0	85	90	468	184	70	83	107	25	108	270	1507
1988	172	114	65	59	8	42	45	98	155	84	38	12	892
1989	141	82	263	105	78	84	91	161	53	12	94	87	1251
1990	7	28	209	0	94	235	19	5	7	16	20	14	654
1991	133	95	0	0	0	0	0	0	0	0	0	0	228
TOT.	4661	6739	7732	6025	6649	6180	8151	9102	7245	5683	6739	3818	78724

Table 3 – Year-month distribution of the oceanographic stations in the Indian Ocean

30°E-80°E / 30°S-30°N

YR	J	F	M	A	M	J	J	A	S	O	N	D	TOTAL
1906	0	0	0	0	1	4	0	0	0	0	0	0	5
1920	0	0	0	0	0	22	30	28	0	0	0	17	97
1921	14	10	0	0	0	17	29	34	19	0	0	0	123
1923	1	0	0	0	11	9	0	0	0	29	38	72	160
1924	12	14	95	17	19	1	0	0	0	0	0	0	158
1928	0	0	0	0	0	0	1	0	1	1	0	1	4
1929	0	0	1	7	2	0	0	4	2	0	1	9	26
1930	11	0	0	0	0	0	0	0	2	0	0	0	13
1932	0	0	0	0	0	3	0	0	0	0	0	0	3
1933	0	0	0	0	0	0	0	0	13	18	19	16	66
1934	4	13	3	7	20	0	0	0	0	0	0	0	47
1935	0	0	0	9	4	0	0	0	0	0	0	0	13
1948	0	0	2	12	73	4	5	0	0	0	0	0	96
1949	16	18	0	0	3	0	0	0	0	0	0	8	45
1950	17	6	29	16	4	9	4	6	0	2	2	0	95
1951	0	12	12	10	0	0	0	2	3	0	23	0	62
1952	0	14	0	18	0	0	0	0	0	0	0	0	32
1955	0	0	0	0	2	0	0	0	0	0	11	0	13
1956	2	9	2	0	9	12	0	0	0	0	0	12	46
1957	2	0	9	0	0	0	0	0	32	78	113	44	278
1958	1	16	10	66	14	44	5	5	24	20	17	9	231
1959	18	34	17	41	31	16	3	4	38	4	14	28	248
1960	25	90	93	45	0	5	26	47	29	0	0	4	364
1961	137	50	74	54	80	21	7	2	2	2	8	2	439
1962	4	2	6	5	10	5	42	73	125	110	78	74	534
1963	127	39	40	15	36	65	109	121	81	89	50	61	833
1964	46	35	74	101	65	74	48	169	73	20	48	53	806
1965	51	59	123	188	150	100	54	19	40	2	4	0	790
1966	56	15	0	0	33	30	51	40	84	64	46	17	436
1967	2	63	20	3	59	19	15	50	39	0	0	0	270
1968	31	71	70	45	25	18	100	198	65	0	73	39	735
1969	23	72	8	37	8	66	0	0	15	89	2	0	320
1970	31	16	18	29	23	79	73	104	101	39	147	52	712
1971	4	92	34	0	96	73	31	41	23	13	71	6	484
1972	15	39	73	80	20	66	75	79	35	23	14	59	578
1973	30	58	92	21	150	116	76	9	9	0	11	10	582
1974	0	41	40	13	38	28	14	3	0	10	7	11	205
1975	42	19	35	12	51	3	10	31	0	1	0	0	204
1976	0	0	1	6	65	18	26	11	0	7	6	0	140
1977	33	54	22	6	0	2	22	28	15	35	15	11	243
1978	11	0	0	18	0	17	0	0	0	0	0	0	46
1979	0	38	8	132	449	250	32	0	0	46	15	1	971
1980	0	7	0	0	48	0	0	37	61	0	7	0	160
1981	4	5	0	7	0	0	0	0	0	7	0	0	23
1982	4	20	0	0	0	27	23	6	29	56	0	20	185
1983	0	0	14	20	0	0	18	64	0	0	0	0	116
1984	0	0	0	0	0	0	0	0	0	1	0	0	1
1985	0	49	34	192	0	0	0	34	0	1	0	0	310
1986	0	1	3	78	3	0	36	12	0	0	0	79	212
1987	126	0	1	0	0	0	129	91	15	58	0	0	420
1988	11	4	0	0	0	0	0	52	4	0	1	2	74
1990	0	0	0	0	0	0	0	0	1	0	0	0	1
1993	0	0	0	0	56	0	0	0	0	0	0	0	56
1994	0	4	14	0	1	0	0	0	0	0	61	0	80
TOTAL	911	1089	1077	1310	1659	1223	1094	1404	980	825	902	717	13191

Table 4 – Year-month distribution of the temperature profiles in the Atlantic Ocean

80°W - 20°E / 30°N - 30°S

YR	J	F	M	A	M	J	J	A	S	O	N	D	TOTAL
1985	309	512	780	538	541	378	408	612	428	518	379	317	5720
1986	603	385	370	433	273	318	351	288	327	416	547	131	4442
1987	342	548	533	359	377	253	320	389	418	240	407	392	4578
1988	270	302	207	185	192	151	294	297	333	380	265	254	3130
1989	285	439	530	359	314	260	271	415	308	216	142	280	3819
1990	253	646	553	367	372	896	528	430	601	690	745	501	6582
1991	689	628	553	334	337	430	377	279	341	459	476	447	5350
1992	396	342	496	565	688	465	347	373	438	302	303	348	5063
1993	389	416	402	390	224	324	406	356	347	308	428	348	4338
1994	360	479	429	314	538	427	502	418	342	463	361	354	4987
1995	402	280	413	391	291	376	323	373	265	371	315	173	3973
1996	260	181	380	295	215	174	149	200	294	247	215	214	2824
1997	285	210	390	399	332	427	306	404	507	490	460	595	4805
1998	53	61	84	105	107	99	29	54	109	101	158	99	1059
1999	107	80	46	72	32	141	56	31	69	41	69	3	747
2000	68	32	35	38	40	32	3	39	44	61	55	1	448
TOT.	5071	5541	6201	5144	4873	5151	4670	4958	5171	5303	5325	4457	61865

Table 5 – Year-monthly distribution of the temperature profiles in the Indian Ocean

20°E - 120°E / 65°S - 30°N

YR	J	F	M	A	M	J	J	A	S	O	N	D	TOTAL
1906	0	0	0	3	2	3	0	0	0	0	0	0	8
1917	1	0	0	0	0	0	0	0	0	0	0	0	1
1926	0	0	7	0	0	0	0	0	0	0	0	0	7
1929	0	0	0	2	1	0	0	0	0	0	1	13	17
1947	0	60	14	0	0	0	0	0	0	0	6	0	80
1948	0	0	2	11	6	0	0	6	0	0	0	6	31
1949	0	0	0	6	2	6	6	0	0	0	6	0	26
1950	0	6	2	3	2	6	0	7	21	5	0	0	52
1951	24	15	0	1	6	6	8	10	3	3	2	0	78
1952	0	0	0	0	0	0	0	0	0	0	2	0	2
1954	3	3	2	3	1	3	2	3	9	3	2	2	36
1955	2	8	3	1	6	2	7	2	3	2	10	2	48
1956	11	13	2	1	34	11	2	4	3	4	4	7	96
1957	12	12	59	44	7	10	7	0	6	8	28	5	198
1958	13	17	37	151	40	2	2	34	3	11	2	6	318
1959	0	11	9	0	12	17	24	0	22	15	28	91	229
1960	356	570	127	13	8	48	32	45	33	15	27	129	1403
1961	46	41	38	133	14	10	35	10	13	26	14	20	400
1962	15	22	47	42	7	129	142	86	58	73	164	27	812
1963	54	13	47	59	70	47	79	30	217	56	72	30	774
1964	46	38	64	108	43	84	60	153	56	27	13	33	725
1965	78	34	52	74	129	135	358	20	13	81	9	69	1052
1966	23	27	66	12	35	33	13	24	77	86	93	80	569
1967	51	117	49	21	18	22	53	58	139	124	125	83	860
1968	108	168	213	99	131	35	61	79	90	218	155	152	1509
1969	261	306	289	102	71	73	94	73	213	172	49	113	1816
1970	273	282	81	93	98	102	90	191	188	233	280	225	2136
1971	171	412	417	150	176	293	299	339	328	297	428	259	3569
1972	384	414	255	218	113	204	235	115	121	168	93	305	2625
1973	329	325	143	6	126	79	117	84	73	86	123	106	1597
1974	511	506	118	23	96	63	76	49	102	218	60	37	1859
1975	274	348	143	52	85	107	30	179	218	177	236	92	1941
1976	244	313	144	17	146	106	25	27	137	204	220	176	1759
1977	280	216	81	0	75	44	0	2	143	167	64	61	1133
1978	199	226	161	41	72	55	12	36	159	193	31	60	1245
1979	241	375	107	26	148	152	34	25	191	262	110	73	1744
1980	137	157	240	142	420	287	327	297	424	337	175	182	3125
1981	340	468	388	454	374	329	210	226	447	615	549	383	4783
1982	360	409	384	280	335	537	292	206	517	541	545	480	4886
1983	372	383	197	235	372	343	138	276	387	669	310	258	3940
1984	341	448	243	171	354	271	212	256	533	519	263	166	3777
1985	462	475	305	280	163	425	176	225	567	680	385	256	4399
1986	249	167	101	214	89	212	166	74	330	635	349	188	2774
1987	288	245	370	198	172	122	220	254	384	378	238	122	2991
1988	189	248	179	225	188	279	148	185	186	404	293	292	2816
1989	259	313	251	305	430	316	434	249	399	441	396	256	4049
1990	291	321	296	406	307	382	323	404	353	500	539	469	4591
1991	335	380	415	365	420	484	329	245	325	318	467	362	4445
1992	320	465	298	234	257	136	277	318	212	513	415	353	3798
1993	400	498	632	383	717	437	599	423	709	611	800	570	6779
1994	560	535	663	667	483	290	316	529	325	448	545	411	5772
1995	323	389	364	545	465	248	254	722	660	654	563	631	5818
1996	434	694	1022	686	494	321	296	402	328	445	335	165	5622
1997	91	84	17	39	74	60	50	54	57	59	76	66	727
1998	39	36	48	64	52	32	54	33	46	78	52	50	584
1999	81	33	103	43	63	61	45	67	46	30	23	40	635
2000	9	0	31	0	0	19	3	0	26	0	0	0	88
TOT.	9890	11646	9326	7451	8009	7478	6772	7136	9900	11809	9775	7962	107154

Figure 1 – Quarterly distribution of oceanographic stations in the Atlantic Ocean (1906-1991)

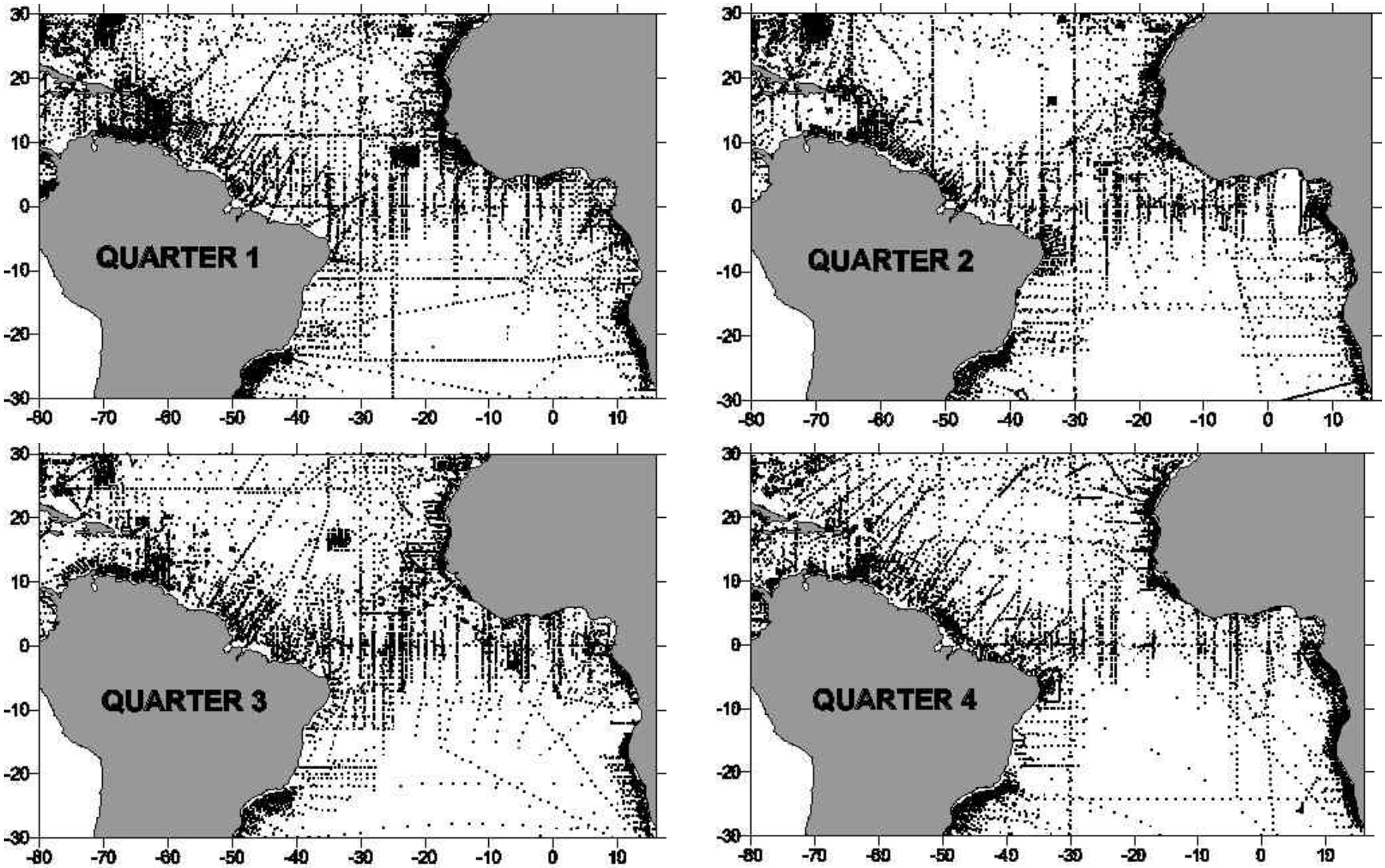

Figure 2 – Overall distribution of oceanographic stations in the Indian Ocean (1906-1994)

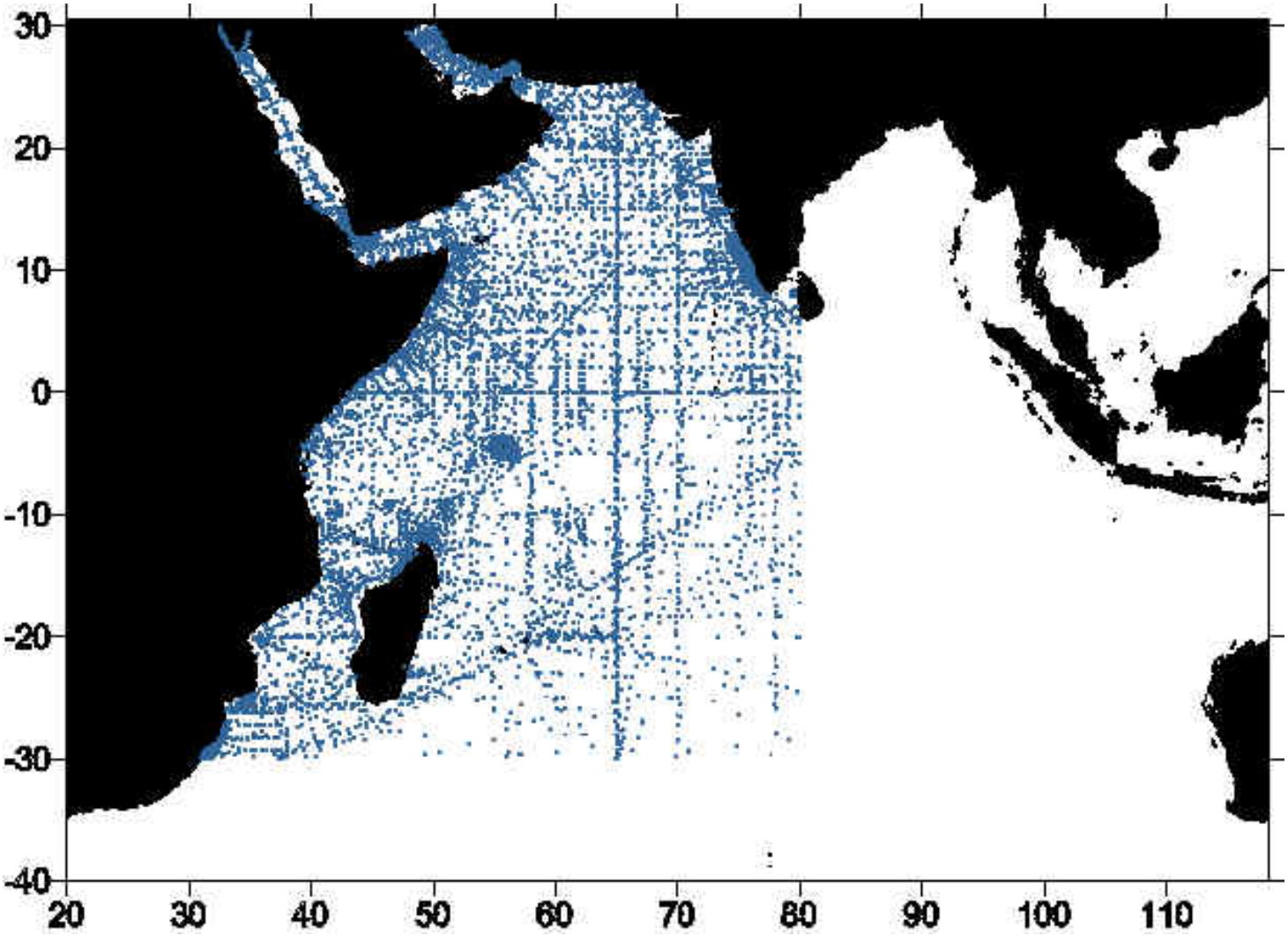

Figure 3 – Quarterly distribution of temperature profiles in the Atlantic Ocean (1985-2000)

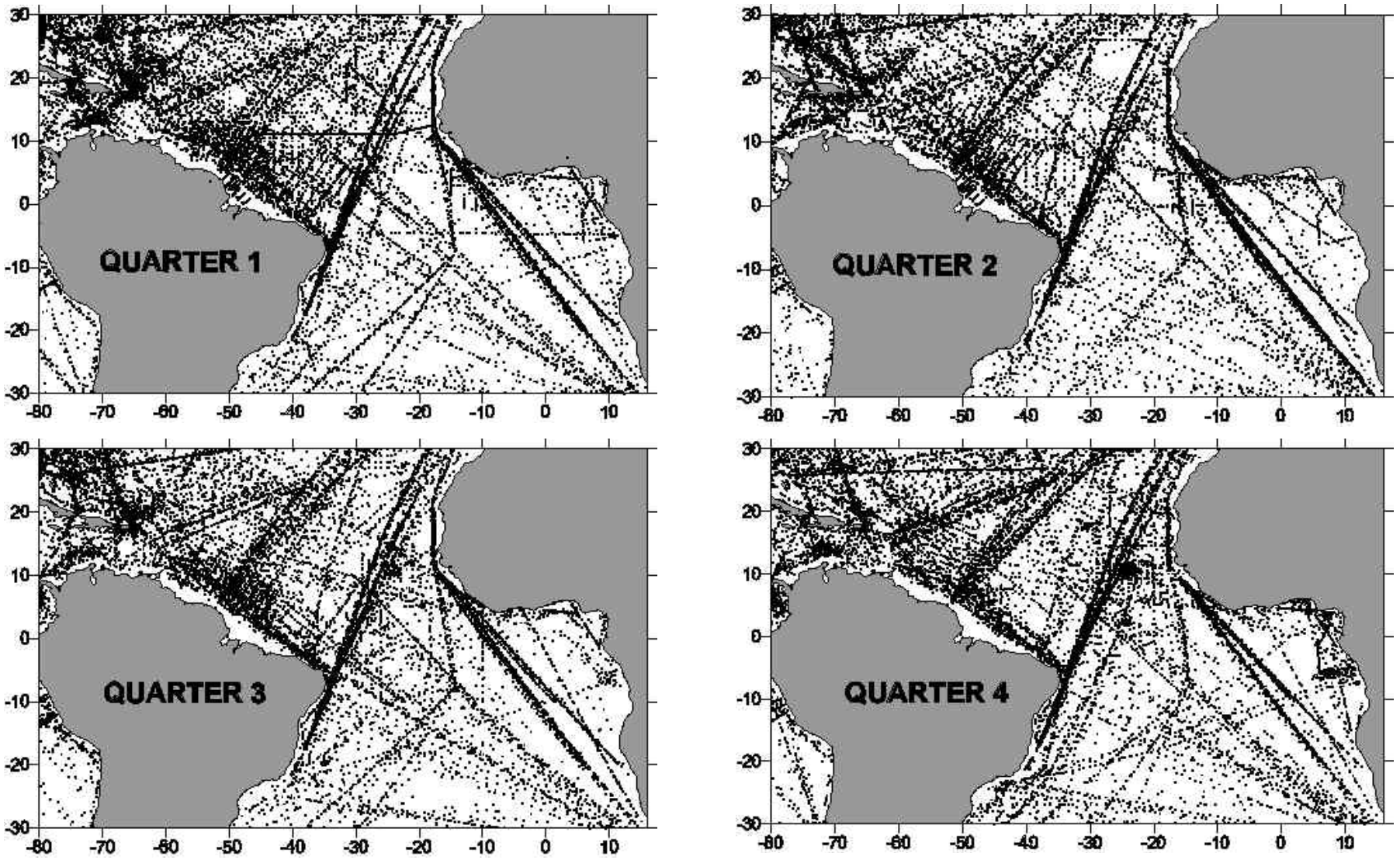

Figure 4 – Quarterly distribution of temperature profiles in the Indian Ocean (1906-2000)