Routines for creating your own mirror of the Argo data files

Last update December 12th 2008

Let us know if you use this routine: Send an email to Howard.Freeland AT dfo-mpo.gc.ca (but replace "AT" with the @ symbol) if you use this routine, I will add your name to a mailing list and let you know if there is any update, change or improvement to the routine.

Change Notes:

(1) December 12th 2008: The US GDAC has moved to a new server. As the address of the ftp servers is hard-wired into the PERL script (driver.pl) the script has been modified by Howard Freeland and appears to be running smoothly.

(2) May 1st 2006: An error in how the Profile Quality flags were extracted has been identified and fixed. It is <u>highly</u> recommended that users first upgrade to the new version by download the zip file, see below, and extracting the files inside, then delete all files under "pacific_ocean", "atlantic_ocean" and "indian_ocean" then do a new download 2001 to present. This problem resulted from the recent format change at the GDACs.

(3) November 1st 2005: The program has been changed to be more efficient about not re-doing things that have previously been done, and so runs much faster. Also, the date entry has been changed to allow the user to select specific months to update. This update is strongly recommended. There are no changes in the resulting data, the entire process is just more efficient and faster. Please download driver.zip (see below) and after extracting the files inside it, read the release notes.

Click here to return to Howard's Argo pages.

Many people have difficulty working with the netCDF format of files on the global Argo data servers. To encourage wider use of Argo data we have created some software that will create an ascii image of the Argo data files on your own computer. But beware, this can take a lot of time.

Requirements:

1) Make sure that you have a fast computer with plenty of space on the hard drive.
2) Most importantly, this program will download a large amount of information by an automated ftp process. Ensure that you have a fast internet connection. It does not make sense to run these routines through a telephone modem.
3) At the moment this program runs only on PCs under Windows. We were working on a Unix version, but have given this up for the time being.

Procedure:

1) Download a file called "driver.zip" into an empty directory. This will become the root directory of your Argo archive.

if you want to start the download (922 kBytes).
2) Unzip the the zip file in place.
3) Run the file driver.bat from the DOS prompt at some time when you can leave the computer alone for some time. There are many options for running driver.bat. Typing the word Help in the DOS window will summarise all options, but before outlining those options it would be best to summarise what the routine does.

Operation:
The standard way of operating the routine is to mirror everything on the Global Argo data servers and the first run should do exactly that, mirror everything. I suggest however limiting the mirror job to the time period 2001 to present. The first time you do this will likely take up to 8 hours to create a base mirror site, tranlated to ASCII and indexed.

The global Argo Data servers are in the USA and in France, two of them carrying identical versions of the Argo data set. You can choose to mirror one or the other, you can use the French server one day and the US server the next day, that is not a problem. The two servers are:-
US Server: ftp://usgodae1.usgodae.org/pub/outgoing/argo/
French Server: ftp://ftp.ifremer.fr/ifremer/argo/
At each of the sites is a geographic approach to archiving profile data from Argo, follow the link "geo", and then you see 3 oceans listed, Indian, Atlantic and Pacific Oceans. Under each of these are years for which Argo-like data exist, under the years 2001, 2002....2005 are 12 directories listing months 1 through 12 and inside each of these directories are files with names like 20020313_prof.nc, which can be read as profile data in netCDF format for the 13th day of March 2002. There will be many profiles inside a single file.

The mirror is organised slightly differently. Underneath your mirror directory (i.e. the directory where you put driver.zip and expanded it) you will create three directories for each of the oceans, as on the Global Data Servers. The routine will also create a directory called "meta" which will carry trajectory and velocity information. Under each ocean directory there will be directories for each of the years, 2001, 2002....2005 etc. Under each year there will be 12 directories representing each month of that year, then it changes. Under each month will be a series of sub-directories representing each day of the month. Each of those directories will contain a series of files, one for each profile acquired. Each profile file is in a flat ASCII format, specifically, a self-describing format known as "IOS Header Format". The format was created by Joe Linguanti, LinguantiJ@pac.dfo-mpo.gc.ca, at the Institute of Ocean Sciences. He can supply manuals desciding the format and has software for looking at IOS Header format data. In addition every one of the daily files contains a file that is always called Index.csv. This is an index file, as its name suggests, inspection of this file will tell you how many profiles were acquired on that day in whatever ocean you are examining, will list the file names, latitude and longitude of the profile and the maximum depth of the profile.

After the initial run and creating your local master copy of the Global Argo Data set, I suggest setting a run of the driver.bat file to run as a Windows scheduled program,

perhaps once per week. When the program re-runs it will examine each netCDF file to see if it has changed, if it has changed then the new version will be updated. If a new netCDF file is downloaded then it will be re-translated into the IOS Header files and re-indexed. New files change as some agencies are slow reporting new data, but very old files can change as data are recalibrated.

Options on running Driver.bat:
One can always get a quick summary of the available options by typing "help" from the DOS prompt.

The instruction format is:-
driver CC Year1/MM Year2/mm O A

CC is a country code, acceptable enties are US us Us FR fr Fr and indicate which of the Global Argo Data Servers you wish to use.

Year1 is the starting year for download, usually for Argo data this will be 2001. But if you want a quick update of a particular year you might specify another year. Inserting a hyphen - will download all available data.

Year2 is the end year for download. Usually this will be the current year but might be something different for a special purpose. Inserting a hyphen - will default to the current year.

MM and mm are month numbers that are optional. If /MM is not included then the program will assume that it is to start at month #1 (January) if /mm is not included then the program assumes that it will download and convert files up to and including the current month.

O is an Ocean code as follows:-
If O = 1 then data for the Atlantic Ocean only will be mirrored.
If O = 2 then data for the Indian Ocean only will be mirrored.
If O = 3 then data for the Indian AND Atlantic Oceans will be mirrored.
If O = 4 then data for the Pacific Ocean only will be mirrored.
If O = 5 then data for the Atlantic and Pacific Oceans will be mirrored.
If O = 6 then data for the Indian and Pacific Oceans will be mirrored.
If O = 7 then data from the Atlantic, Indian and Pacific Oceans will be mirrored
The Ocean argument is required

A is an action code, acceptable entries are P, p, T, t, A or a. These mean:-
P or p - Transfer and process profile information only
T or t - Transfer trajectory information only
A or a - Transfer All information, i.e. both profiles and trajectories.

Examples
driver fr 2001 - 7 A

This will complete a run or update of all data 2001 to the present time, for all three oceans using the french server, profiles and trajectories.

driver us 2003 2003 4 P
This will complete a run updating profile data only for 2003 in the Pacific Ocean using the US Argo data server.

driver fr - - 7 a
This will download everything available at the French global data server.

driver us 2005/9 2005/10 4 P
This will download all Pacific profiles from the US server for September and October 2005 only.

driver us 2005/6 - 4 p
This will download Pacific Ocean profiles from the US server from June 2005 to the present time.

Comments:
A new run of the form "driver fr - - 7 p" may take 24 to 48 hours to run. However a weekly update will likely take only 30 minutes or so.

Doing this over a telephone modem may not be a good idea, depends on your patience and how you pay for telephone access.

The routine does produce error messages. A lot, if not all, of these are trivial and unavoidable annoyances I recommend that you ignore them.

It is a really good idea periodically to do a complete update. It is faster than the initial creation of the mirror because only files that have changed will be downloaded, expanded into ASCII files or Indexed. But if data have been adjusted for data quality issues then this WILL be reflected in your personal mirror.

Trajectory downloads can be time consuming because every time a float reports the entire trajectory file must be altered. In contrast, if we do profile updates we may find that nothing in 2001 or 2002 has changed, so that can be quick. The trajectory files include computations of deep drift velocity. I recommend that you use these instead of trying to compute your own as there are tricks to estimate the actual time at the surface etc.

Note that in each profile data file there are two columns of pressure, two columns of salinity and two columns of temperature. Always read the second column. Frequently the two columns are identical, but if data are changed for quality assurance then it is always the second column that will change. The first column will always carry the raw data.

Pay attention to the quality flags. If you want to know what those mean then log onto one of the data servers and read the manual. The Argo data manuals are available from here...