

Quality control using QCed and Mquest: A Comparison

Nisha Kurian

[National Institute of Oceanography, Goa, India]

During my stay at OCL, NODC from October 10, 2010 to January 10, 2011, I have compared the features of two softwares developed by the United States National Oceanographic Data Center (NODC) and Australian Commonwealth Scientific and Research Organization (CSIRO) for the quality control (QC) of the temperature data. A total of 3300 temperature profiles were quality checked using the NODC developed Global Temperature Salinity Profile Programme (GTSP) QC software and the CSIRO developed Mquest.

Any dataset is required to undergo scientific quality control before it is taken up for analysis. This validation of the data can be done by combining statistical analysis with the climatology and the regional behavior of the parameter undergoing the quality check. Such a validation of the data helps in identifying the real feature against the error in the data due to malfunction or failure of the instrument used. This is important since a high quality data alone can provide details on the variability of oceanic features over different time scales.

A visual approach of the data is available in both the softwares. Each profile is checked against climatology for possible realistic oceanic features or the errors and a quality class of 0-5 is assigned based on World Ocean Climate Experiment (WOCE) definition of the data. [Table I]

NODC QCed software:

The NODC GTSP QC editor (QCed) is IDL-based software which can be run using the IDL "Virtual Machine" (the full IDL installation is not required). It is available for download with the installation instructions at <http://www.nodc.noaa.gov/GTSP/tools/index.html>. The website also contains a perl script which will assist the user to read and write GTSP ASCII format.

Unzip the 'qced.zip' file downloaded from the website to "C:\qced". The file 'qced_install_notes.txt' provides details on installation guidelines. More example data is available for download at <http://www.nodc.noaa.gov/GTSP/gtspp-home.html>. The configuration file may be edited to add additional Action codes. Also contact Melanie Hamilton (Melanie.Hamilton@noaa.gov) in case codes are being added to the data send to NODC.

The graphical user interface (GUI) allows loading of the data for visual QC editing. The GUI has a map, focus controls, status and trouble lights, display and edit options and a scrolling list of the stations. The map has on top of it the full path name of current data set displayed. The map shows the position of the stations in the world oceans. The currently selected station is shown in red, other stations in same cruise (same call sign or plat id) is in orange while the stations from other cruises in the file are shown in green (Figure 1). The focus control helps to shift the focus from current profile to other profiles based on next station, next flagged station etc. During loading or saving of MEDS-ASCII file the status and trouble lights indicates the percentage of task completed whereas it indicates the status of error flags of the current profile during other times. The various checks applied in GTSP software are available at <http://www.nodc.noaa.gov/GTSPP/document/process/qcproces.html>.

The scientific QC of the profiles is carried out as follows. To edit a profile, it has to be brought up as a profile plot. Each individual profile is plotted against a background monthly climatology of the corresponding month, seasonal mean and annual mean. Each individual profile has three colored lines indicating the QC flags associated.(Figure 2) The first columns of colored dots shows the original (incoming file) setting for the flags. The middle column shows what those flags look like after the automatic tests have been performed (and how they will be recorded in the output file if the operator does no further editing). The right-most column shows the flags after the operator has done any hand-editing, and how they will be recorded in the output file. The waterfall plot displays up to fifteen neighboring stations at once for side by side comparisons. All the stations displayed are from the current cruise where the current station is displayed in red and the other stations in black (Figure 3). This helps in checking the consistency of the profile with respect to the neighboring profiles.

CSIRO Mquest:

Mquest is a MATLAB based QC software. Since a compiled version of Mquest is available at <ftp://ftp.marine.csiro.au/pub/cowley>, it is not necessary to have MATLAB installed on the PC for working with Mquest. The website contains two files, Mquest_pkg.exe and Mquest_files.exe. The program Mquest_pkg.exe is a self-inflating installer which will install the matlab runtime programs it needs and the executable is called 'Mquest.exe'. These files from Mquest.exe should be extracted to “:C:\Program Files\Mquest”. Also, to run the program, download files from http://www.marine.csiro.au/~gronell/Mquest/index_files/page0001.htm.

Unzip the files downloaded and put them in directories as follows:

terrainbase.nc.gz “C:\Program Files\Mquest\terrainbase.nc”

CARSatlas.tar.gz “C:\Program Files\Mquest\CARSatlas”
levitusdatanew.tar.gz” C:\Program Files\Mquest\levitusdata”

The Mquest.exe file and the files to be QC’ed should be present in the same directory. Before Mquest can access the files, the data should be in the Mquest NetCDF (MQNC) format. In this format a single netcdf file can hold multiple profiles (eg, temperature, salinity and conductivity) of each station but, each station is held in separate files. This format is a hybrid of the Argo format and the WOCE format used by NODC (for format details of the data see Appendix B of Mquest manual available at <http://www.marine.csiro.au/~gronell/Mquest/MQuESTmanual.pdf>)

The underlying file structure is based on the MEDS-ASCII GTSP format used by NODC. However, the file extension needs to be .MA unlike the .meds used by the NODC. The file was renamed (XXXX.meds to XXXX.MA) and it was made sure that no spaces or underscores are present in the filename. Each database is structured as a directory tree with a name or ‘prefix’ identifying the top directory, with subdirectories that lead to the profile files. The creation of this directory tree is taken care when the database is created. Each directory contains at most 200 profiles (0-99), 100 each with a suffix of “ed.nc” (edited data) and “raw.nc” (raw, untouched data). The files are found using a unique number associated with the data. All the metadata required to retrieve any profile from database is kept in a netcdf file-the keys file. The keys file associated with each data base is named ‘databaseprefix_keys.nc’ and includes the variables such as latitude, longitude, date, time, unique ID or station number, ship callsign, data type and whether the profile has failed one or more of the statistical screening. Mquest allows selecting the profiles based on month or year of the profile and by callsign, time and date or latitude before processing begins. The instructions on how the data should be sorted for QC and the subset to be viewed (month/year) and the files to be used as buddies are to be given when Mquest starts (Figure 4). The buddies are the databases that will be displayed along with the profile being QC’ed to help identify a feature observed as real or a data error.

The GUI has the map, waterfall profiles, scrolling list, the depth-temperature data of current profile, the current profile display, the accept and reject buttons and the details of the current profile like the position, ship name, data type, data source etc. The map has the station locations in the world oceans with the current station in green, the ones that are already examined in yellow and rest of them in blue. The waterfall profiles shows 25 profiles and the current profile is shown in red, QC’ed profiles are in green and the remaining ones are in blue (Figure 5). If some part of the data is rejected while QC’ing, that part of the profile remains red when we move to the next profile. This is a good check to know if all the bad

data is removed. The current profile is plotted against 3-sigma error band of the climatology of the region. The climatology used by Mquest is the CSIRO Atlas of Regional Seas (CARS). It is based on World Ocean Database 2005 (WOD05) [July 2008 Update], surface-pressure-corrected Argo global archives to May 2009, WOCE Global Hydrographic Program (v3.0), and many other datasets available up to 2008.

The buddy checking available in Mquest is helpful in deciding whether an anomaly observed is real or the result of a malfunction. Figure 6 and 7 shows the same temperature profile plotted using QCed and Mquest respectively. The temperature profile in figure 7 lies outside the 3 standard deviation of the climatology of the region. After comparing with the climatology of the region, the profile is checked against the buddy profiles. The profile in figure 7 is now plotted against 10 closest profiles and the result is shown in figure 8. The profile appears to follow the typical temperature pattern of the region in spite of falling out of the climatology band. Hence the quality flag 'QC' is applied with a data class of 1 or good quality data. Figure 9 shows a temperature profile plotted against buddy profiles from different files. The green and yellow profiles are from the same file. If the current profile is from north of 35 deg S, then the profiles to the east of the current profile are yellow and those from west are in green. If the current profile is south of 35 Deg S, the profiles to the north of current profile are green and those towards the south are in yellow. This gives detailed information on the oceanic structure. For example, buddy profiles from south would be expected to be colder than buddies from the north. The buddy profiles from other files are shown in red and magenta. The magenta profiles are from the same year as the current profile is whereas red is from other years of observation. When buddies are displayed, climatology disappears. "Page up" and "Page down" buttons toggles between the buddy display and single profile. After the QC editing of a file is finished, the data is exported using the export button present in the Mquest window to any of the four formats available as per the requirement/ choice of the editor.

Comparisons

Both the softwares follow the WOCE definition for assigning the quality flags. QCed, as well as Mquest, does a check with the previous and next drop to check for the profile to profile consistency using the waterfall plots. This is a common application for both the QC softwares that helps in identifying the realistic oceanic features against the instrument malfunction. A comparison on the features of Mquest and QCed is provided in table II.

The climatology used to check for the realistic features or errors of a temperature profile is different for Mquest and QCed. When the profiles fall out of climatology for Mquest, the buddy checking showed that the temperature profile was typical for the region for the particular month (Figure 7 & 8). But for the same profile, the climatology used by QCed showed that the profile was within the acceptable ranges of

temperature for the region for the particular month. In majority of cases where the temperature profiles were outside the WOA05 climatology band for QCed, Mquest climatology as well as the buddy checking showed similar results. The automatic checks using the CARS climatology flagged many good XBT drops for further check, while the WOA05 did not, for the Indian Ocean.

Mquest allows the usage of shortcut keyboard keys for flagging the data. For example, if there is a surface spike found, in Mquest the shortcut 'qc' chops the surface spike, adds the QC label and the upward key takes to the next profile. A list of short cut keys useful in entering QC flags and moving within a dataset is given at the Appendix F of the Mquest manual (<http://www.marine.csiro.au/~gronell/Mquest/MQuESTmanual.pdf>). While using QCed, for removing the surface spike, one need to first click at the surface level of the profile, mark it as surface spike from the drop down menu, accept the edit marks, go to the next level (from the depth where the profile is acceptable), click at the level and add FR/FB from dropdown menu, accept flag edits, click the exit button and then answer no to the question of 'do you want to continue editing this profile' and then click for the next profile in the file.

The expanding of the profile at any particular depth is another area where the two softwares differ. It is important at times to have a closer look of the profiles to decide if a feature is to be accepted or not. In such cases, Mquest offers a closer examination by making the selected depth as the centre with 100m above and below. The temperature scale is also re-arranged automatically accordingly instead of maintaining the standard -5 to 35 deg C. The QCed software also provides an option of examining the profile in 100 m depth detail, but does not centre at the point selected (It provides between 0-100,100-200,200-300 and so on). This is not very helpful especially when the feature is close to a depth of hundred or multiple of hundreds. For example, figure 11 shows a profile plotted in both QCed and Mquest. It appears that close to a depth of 200m there is a hit bottom. To confirm this, the profile is expanded at 200m depth in both the softwares and is shown in figure 12. Even after expanding the depth, it is difficult to confirm it as a hit bottom with visual aid in GTSP, whereas the Mquest does a good job in this regard.

The strongest feature in Mquest is the buddy checking of the profiles. This comes as a helpful tool as mentioned above where the profile falls out of climatology, but similar trend is observed in profiles of previous and next drops as well as those come from a selected radius (which also can be changed from 0.1 deg to 4 deg) belonging to the same month of either same year or from different years. The QCed seems to have a similar option of comparing with the neighboring profiles but not yet activated. Activation of this option in GTSP would be highly appreciable for Science QC'ing.

The automatic quality check in Mquest also identifies the repeated profiles when the data has been sorted by ship/date/time and a warning of repeat profile is issued. A repeat profile warning is given when next profile is taken within 15 minutes. Most of the times when a repeat warning was given, the repeat appeared to be done since the first drop was a bad data.

The depth- temperature pair of data that is displayed next to the profile in the Mquest provides ease in identifying the exact location of the features like spikes or wire stretch. QCed also provides a visual access to the values of the depth temperature pair in a separate text box on clicking the text button. However, once the text window is open, the QCed does not allow editing the profile.

Mquest allows the manual change of the data values. For example, when a spike is observed which is only at a particular depth or a smaller depth range, the option “SPA” which is an accept code for spikes provides a pop up window. The pop up window allows the editor to decide the depth range over which the spike will be interpolated. However, the QCed does not really allow the editor to change any data value.

Suggestion:

It would be highly appreciable if there can be an autosave option available in QCed. (Once I lost almost half days work done due to the sudden connection failure with the server. I had almost finished a file of 144 profiles and there was an unexpected problem with the server abalone. I lost all the editing I did till then for that file). The QCed also could include the option of expansion of the profiles at the desired depth for the closer examination of features. The flexibility of the temperature scales according to the chosen depth may also be included. If short cut keys could be provided for navigating through the profiles as well as for flagging them, the time required for QC'ing could be highly reduced.

Table I

Class	Quality	Description
Class 0	No QC Done	Class 0 data is the level at which all data enters the working achieve, and has yet not been quality controlled
Class 1	Good Data	Class 1 data is top quality data in which no malfunctions are identified and all real features have been verified during quality control process
Class 2	“Probably” Good Data	Class 2 data is good data in which some unusual but probably real features which are unconfirmed, and/or malfunction errors which can be corrected or are small enough to be ignored without seriously affecting the overall quality of the data, are observed.
Class 3	“Probably” Bad Data	Class 3 data is suspect data in which some unusual and probably erroneous features are observed.
Class 4	Bad Data	Class 4 data is bad data in which obviously erroneous values are observed.
Class 5	Value changed	Class 5 data indicates that the original data value has been altered.

Table II

Features	Mquest	QCed
Comparison with monthly Climatology	Yes (CARS Climatology)	Yes (WOA 05)
Comparison with seasonal and annual climatology	No	Yes
Profile to profile consistency(waterfall profiles)	Yes	Yes
Buddy/neighboring profiles checking	Yes	No
Key board short cuts	Yes	No
Manual editing of the data value	Yes	No
Manual changing of the data quality indicator	No	Yes
Auto save of the QC completed	Yes	Yes
Excessive speed warning	Yes	Yes
Repeat profile warning	Yes	No
Cruise track visual inspection	Yes	Yes
Choice to edit wrong location/date/time data	Yes	Yes
Continuous display of depth-temperature pair	Yes	No

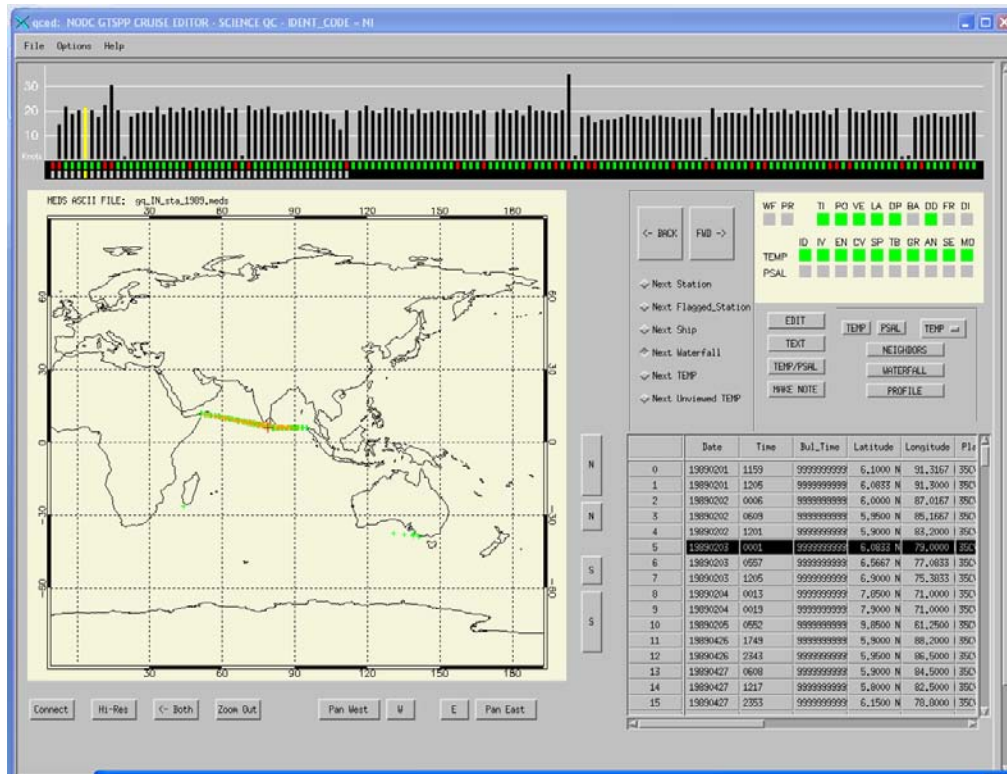


Figure 1: The location of the current profile is shown in red, the profiles from same cruise are in orange and the rest are in green.

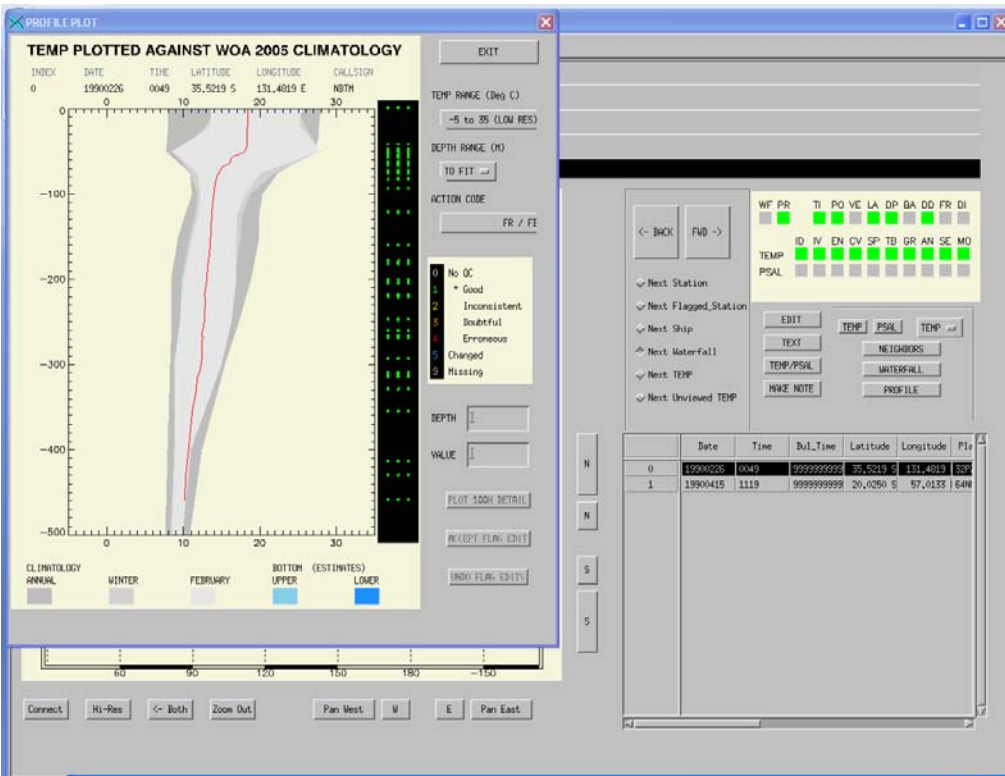


Figure 2: The three lines next to the profile shows the Qc flags associated with first column showing incoming file setting for the flag, middle column showing the results of automatic tests performed and third column showing the editing done by the operator.

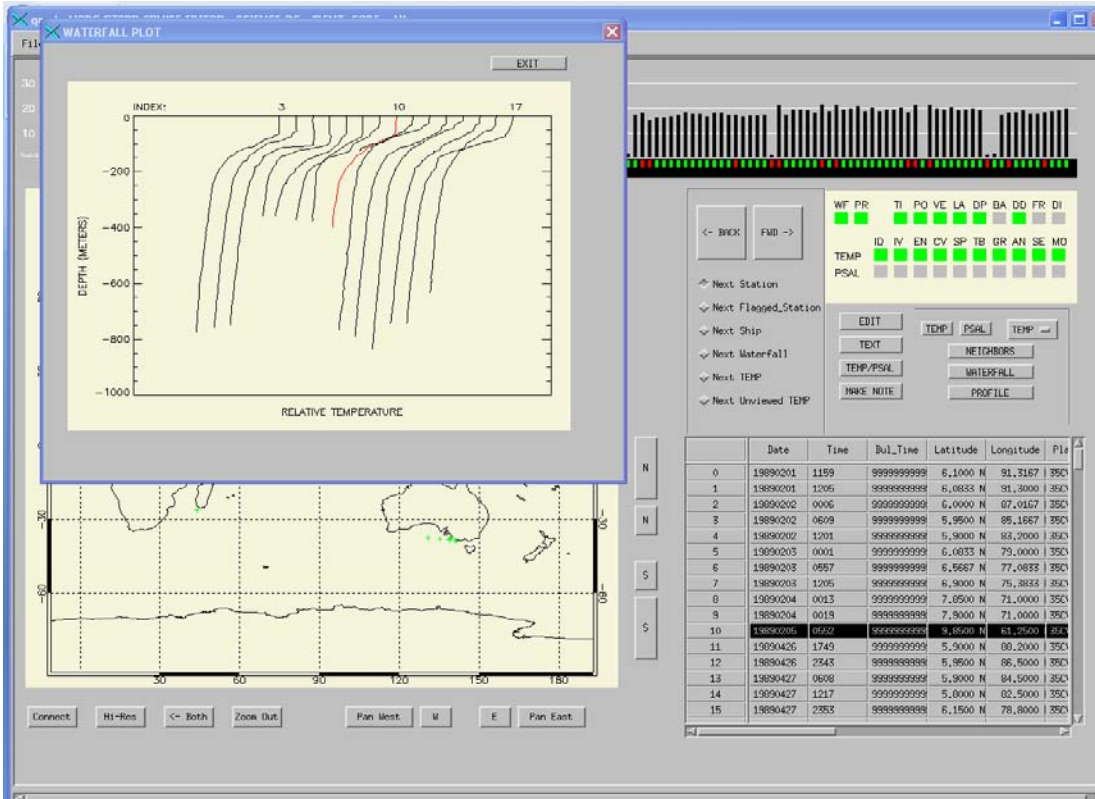


Figure 3: Waterfall profiles in GTSPPP QC

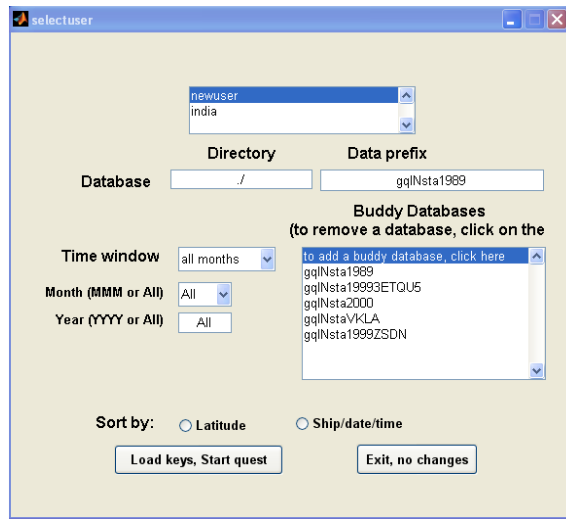


Figure 4: The Mquest window where the data to be edited and the buddy files are assigned. Also the instruction on how the data should be sorted and subsets to be viewed are assigned.

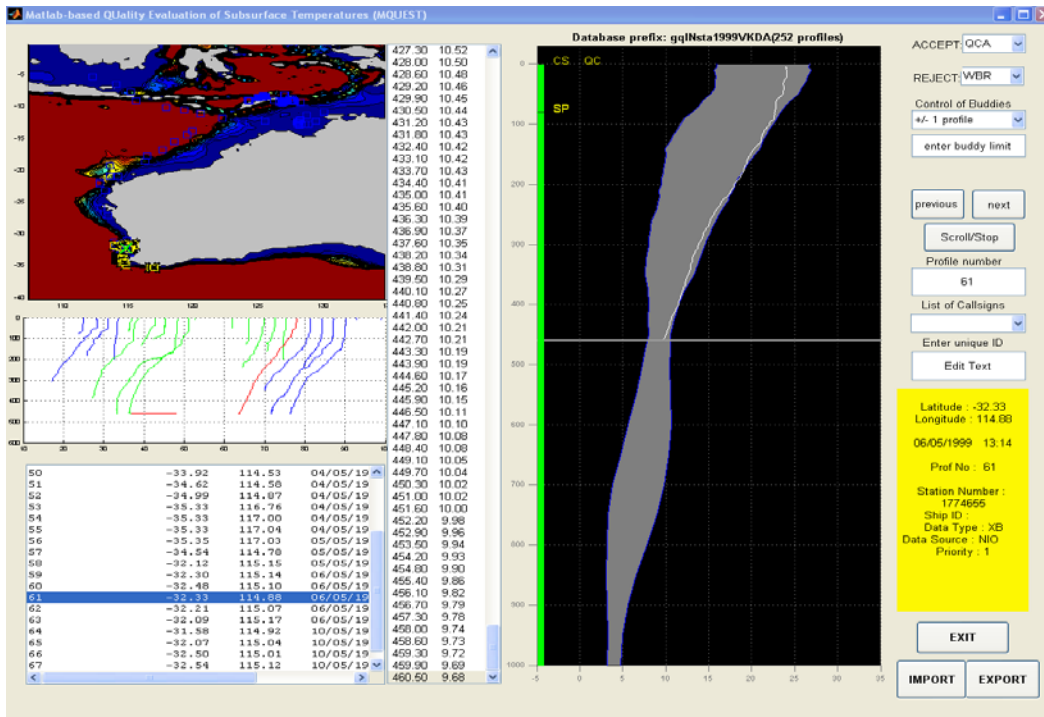


Figure 5: Mquest QC window. The examined profiles are in green current profile in red and the rest of them in blue. The part of the profile discarded as bad data is shown in red. The location of current profile

in the map is in green, the examined ones in yellow and the location of remaining profiles are shown in blue.

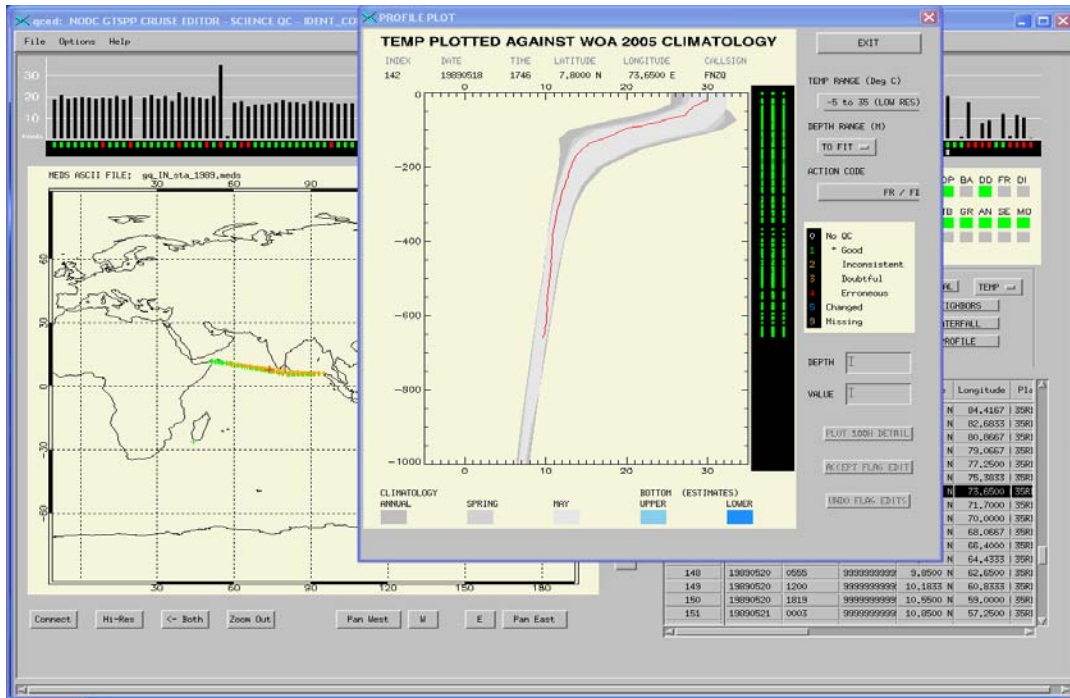


Figure 6: The QC using GTSP. The current profile is plotted against the WOA05 climatology for the region.

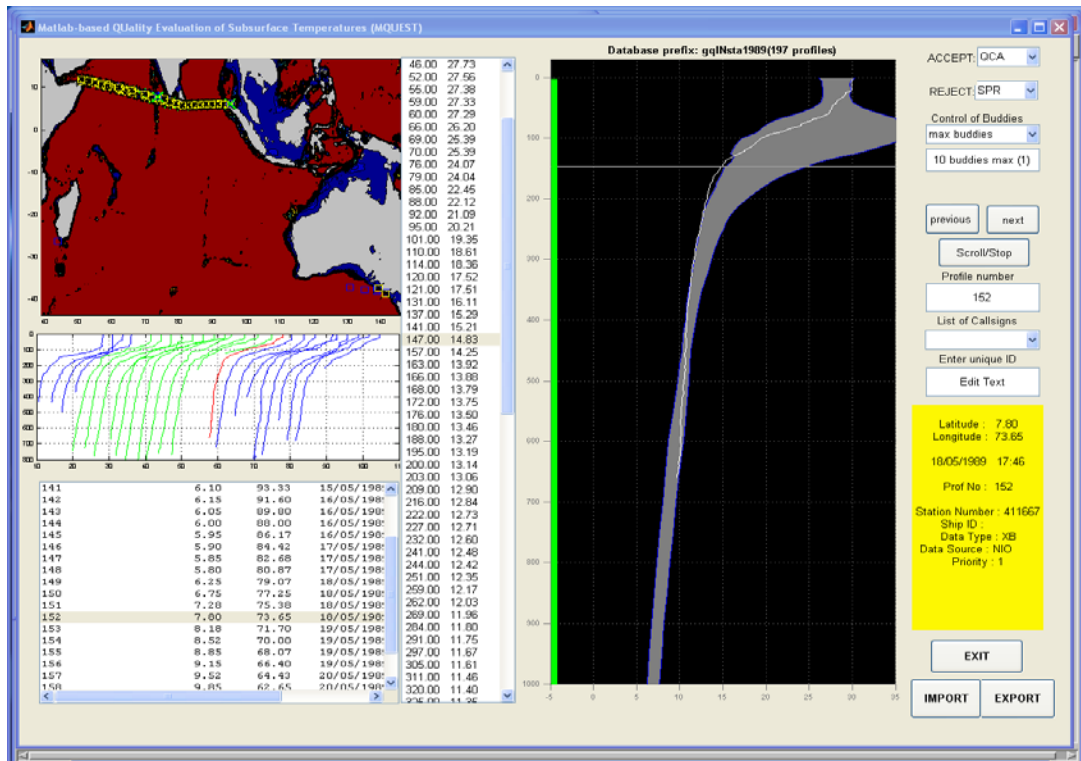


Figure 7: The QC using Mquest, the same profile as in Figure 5 is plotted against CARS climatology

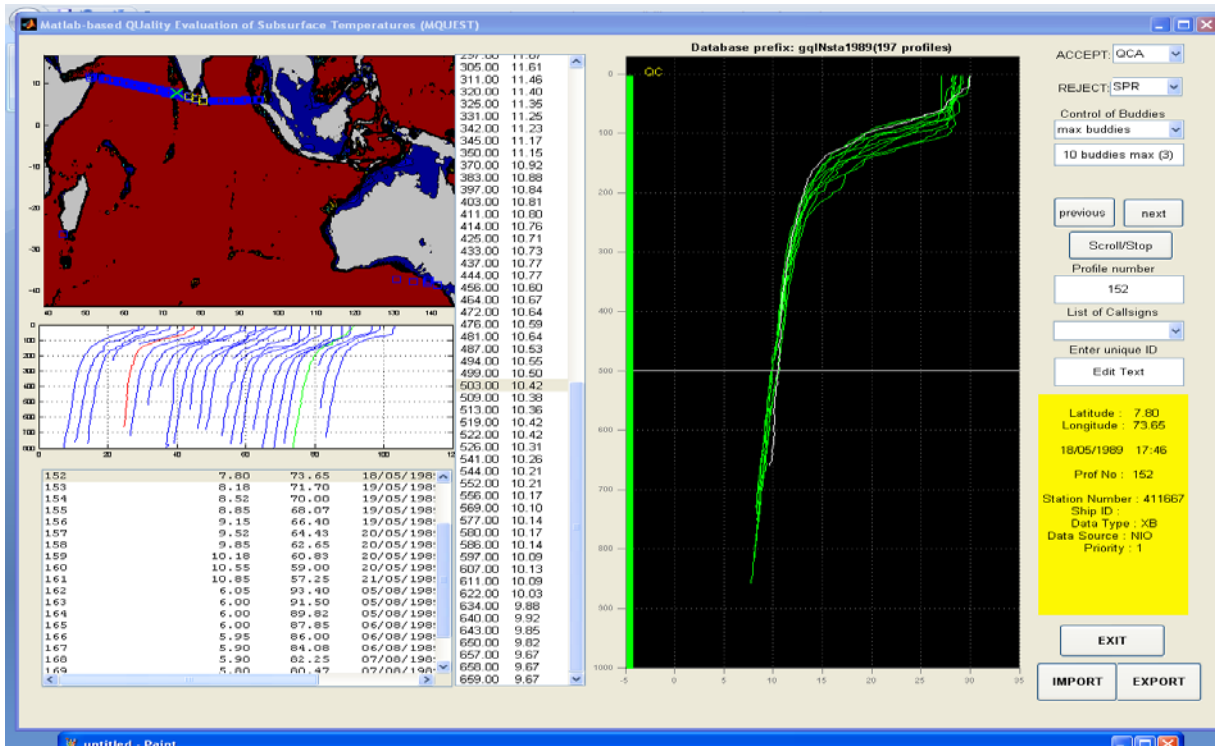


Figure 8: The same profile as in figure 5 and 6, but plotted against buddy profiles.

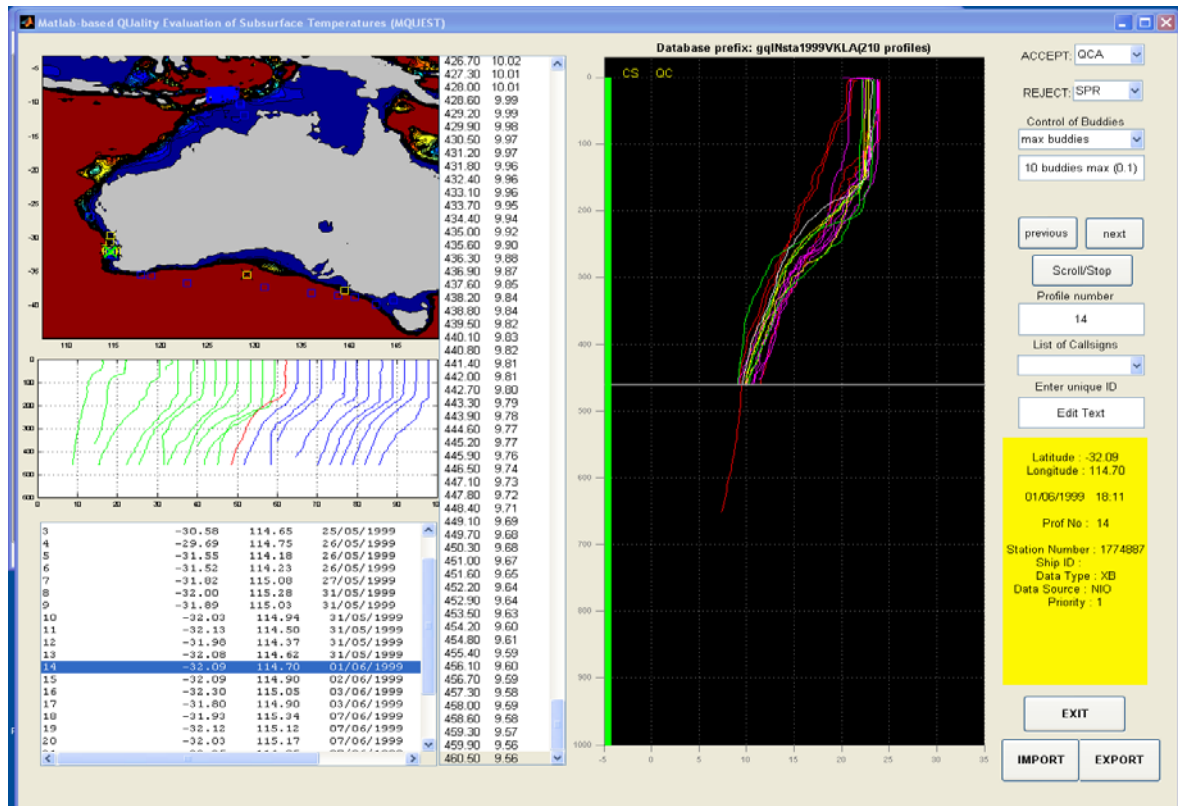


Figure 9: Mquest buddy plotting from different files. Yellow and green are from same files, magenta is from another file but same year. Red from another file and different years.

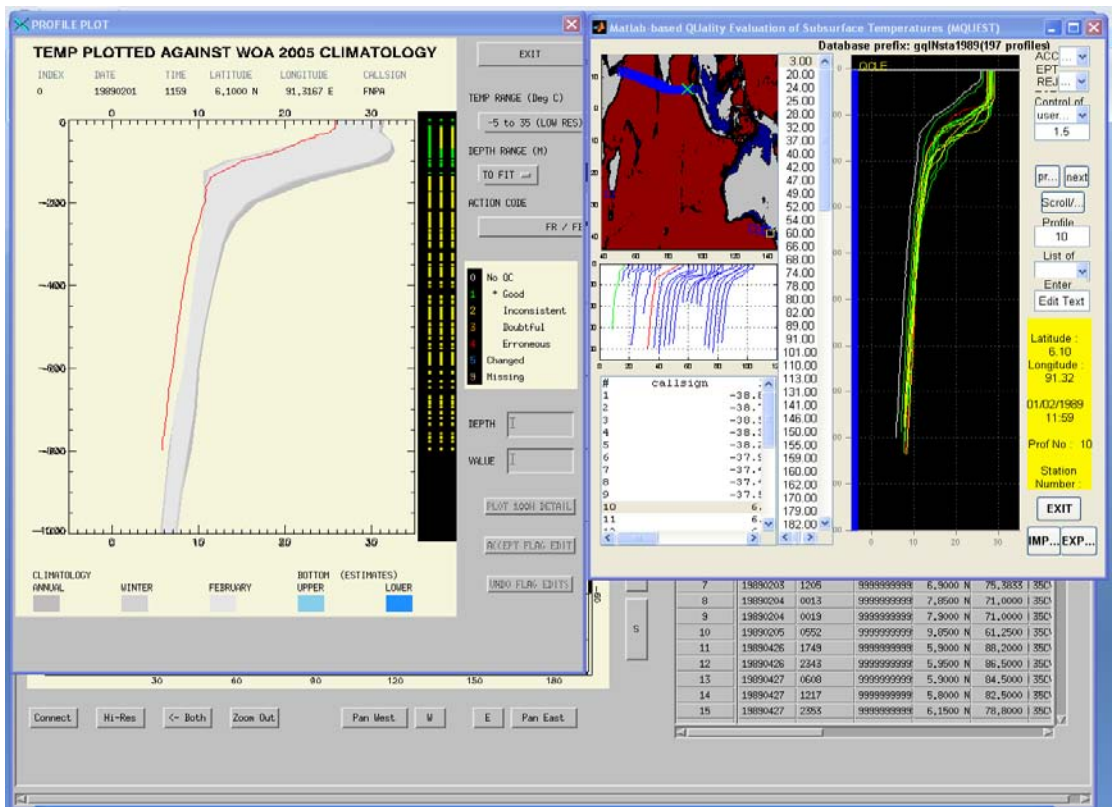


Figure 10: Profile checking using Mquest and GTSP. The profile is outside the WOA 05 climatology and outside the band of buddy profiles.

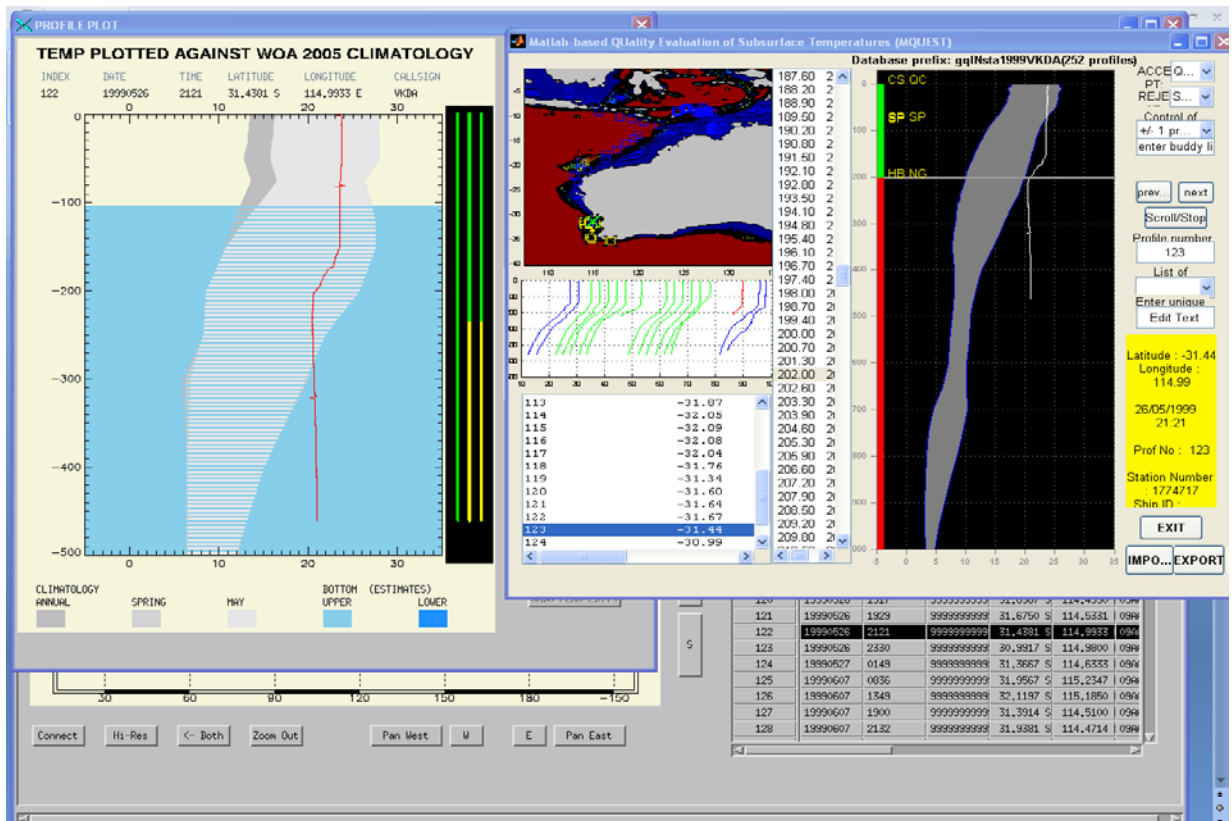


Figure 11: The same profile plotted in GTSP and Mquest for suspected 'hit bottom' close to 200 m.

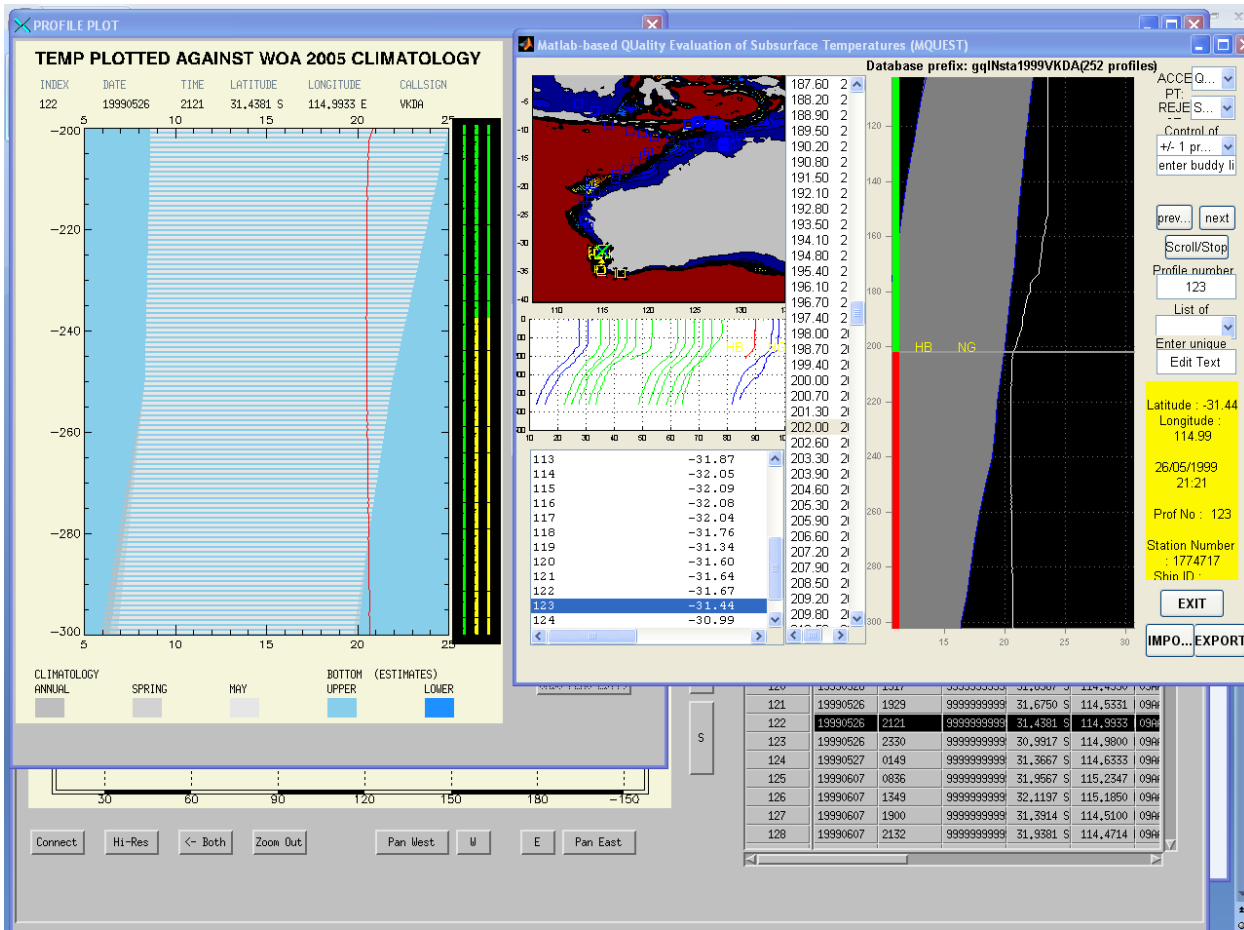


Figure 12: Same as figure 11, but expanded scale close to 200 m.