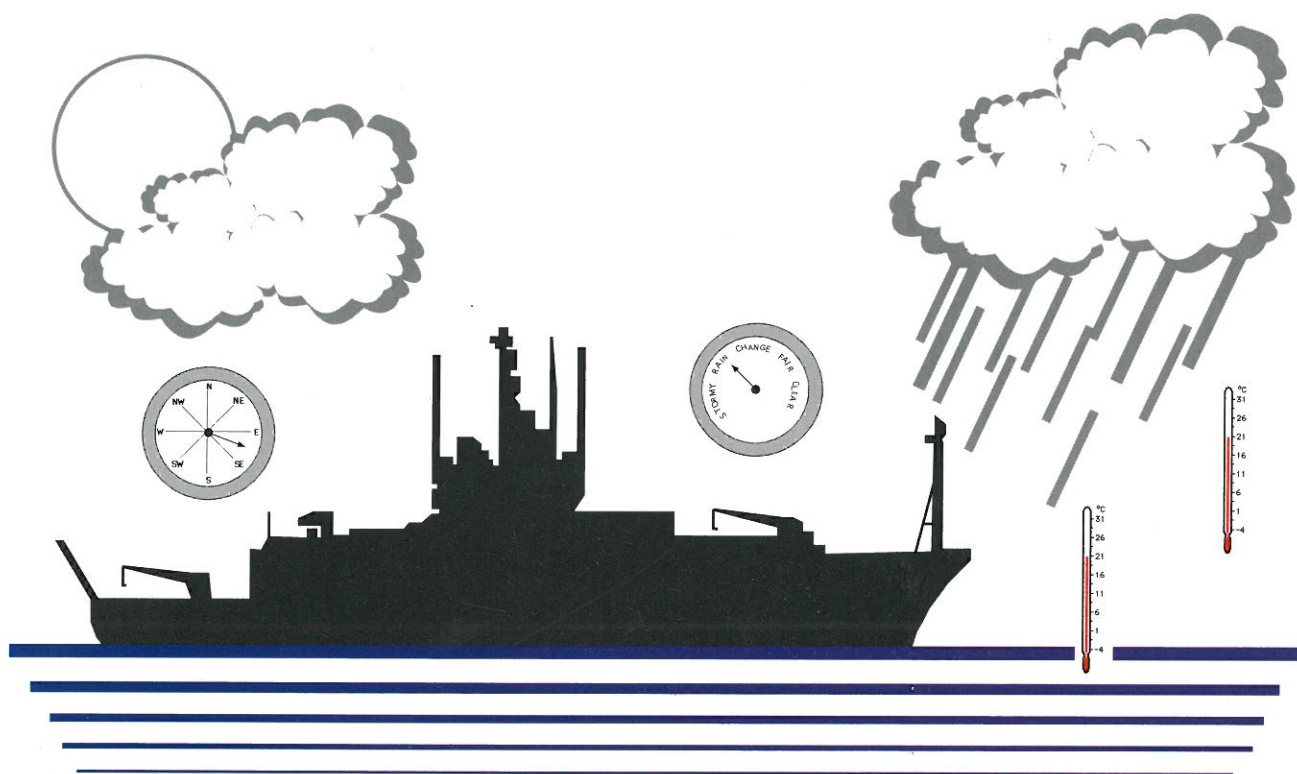
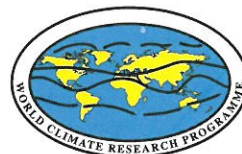


Handbook of Quality Control Procedures and Methods for Surface Meteorology Data



Shawn R. Smith, Christopher Harvey, and David M. Legler

WOCE Data Assembly Center for Surface Meteorology
Center for Ocean Atmospheric Prediction Studies
Florida State University



WOCE Report No. 141/96
COAPS Report No. 96-1

Handbook of Quality Control Procedures and Methods for Surface Meteorology Data

Shawn R. Smith
Christopher Harvey
and
David M. Legler

Data Assembly Center for Surface Meteorology
Center for Ocean Atmospheric Prediction Studies
Florida State University
Tallahassee, FL 32306-3041



WORLD OCEANS CIRCULATION EXPERIMENT
Report No. 141/96

COAPS Technical Report
No. 96-1

15 March 1996

TABLE OF CONTENTS

| | |
|--|-----|
| ACRONYM LIST | ii |
| ACKNOWLEDGEMENTS | iii |
| 1. INTRODUCTION | 1 |
| 2. DATA FORMAT | 5 |
| 2.1 Data conversion to netCDF | 7 |
| 2.2 Flagging philosophy | 9 |
| 2.3 Flags added during data conversion to netCDF | 10 |
| 3. PREPROCESSING | 11 |
| 3.1 Time sequential/duplicate tests | 12 |
| 3.2 Statistical test | 12 |
| 3.3 Bounds test | 14 |
| 3.4 Platform speed test | 16 |
| 3.6 Earth relative wind recomputation test | 20 |
| 3.7 $T \geq T_w \geq T_d$ test | 23 |
| 3.8 Order of precedence for preprocessing flags | 24 |
| 4. VISUAL QUALITY CONTROL PROCEDURES | 25 |
| 4.1 Map window | 27 |
| 4.2 Multiple file plot | 29 |
| 4.3 Editor | 30 |
| Discontinuity | 33 |
| Sensor malfunction | 34 |
| Interesting feature | 34 |
| Erroneous data | 36 |
| Suspect data | 37 |
| Spike in data | 38 |
| Data passing evaluation | 39 |
| 4.4 New file creation and flag documentation | 39 |
| 5. DATA AVAILABILITY | 40 |
| 6. RECOMMENDATIONS FOR FLAG USE | 41 |
| 7. REFERENCES | 42 |
| Appendix A | 43 |
| Appendix B | 45 |
| Appendix C | 49 |

ACRONYM LIST

| | |
|--------|---|
| ASCII | American Standard Code for Information Interchange |
| BODC | British Oceanographic Data Centre |
| COADS | Comprehensive Ocean-Atmosphere Data Set |
| COAPS | Center for Ocean Atmospheric Prediction Studies |
| DAC | Data Assembly Center |
| DQE | Data Quality Evaluator |
| FSU | Florida State University |
| IDL | Interactive Data Language |
| IMET | Improved METeorology ship and buoy system |
| NCAR | National Center for Atmospheric Research (USA) |
| NCDC | National Climate Data Center (USA) |
| NGDC | National Geophysical Data Center (USA) |
| NOAA | National Oceanic and Atmospheric Administration (USA) |
| netCDF | network Common Data Format |
| QC | quality control |
| R/V | research vessel |
| SAC | Special Analysis Center |
| SLP | sea-level pressure |
| UTC | universal time coordinated |
| VIDAT | Visual Data Assessment Tool |
| WOCE | World Ocean Circulation Experiment |
| WMO | World Meteorological Organization |

ACKNOWLEDGEMENTS

COAPS receives its base funding from the Physical Oceanography Section of the Office of Naval Research. Funding for this report was provided by the Physical Oceanography Section of the National Science Foundation, grant number OCE-9314515. The authors wish to acknowledge Mr. Ryan Sharp for the computations in this report. We also wish to thank Dr. Mark A. Bourassa and Dr. Leslie Hartten for their critical comments and assistance in improving the authors understanding of ship wind measurements. The figures were produced with the help of Mr. James Stricherz, Ms. Jiraporn Whalley, and Mr. Parks Camp.

1. INTRODUCTION

The Data Assembly Center (DAC) for Surface Meteorology at Florida State University is charged with collecting, quality controlling, archiving, and distributing all underway surface meteorological data from WOCE vessels worldwide. The scope of the DAC's collection efforts also includes all surface meteorological data from WOCE-sponsored experiments (i.e. Subduction Experiment). Types of data collected include standard ship bridge observations, advanced automated systems (e.g. Improved METeorology (IMET) measurements system), and all other practically obtainable surface meteorological data from WOCE research vessels and buoys.

Once the data are collected, the DAC verifies the accuracy of the measurements using a series of data quality evaluation procedures. The focus of the DAC Handbook of Quality Control Procedures and Methods is to describe the methodology and detail these quality control (QC) procedures. The DAC intends for this document to be used by the WOCE scientific community as well as other interested researchers. The flagging procedures contained in this report may also be of general interest to data evaluators from the international data centers (NCDC, NCAR, BODC, etc.).

The overall goal of the DAC's QC is to provide a well-documented, reliable, and consistent data set to the scientific community. The DAC evaluates the data to the best level possible with the QC results depending primarily on the level of knowledge of the incoming data as well as examination and analysis of the data. Information provided about how the data were collected, i.e. metadata, is an essential component of any data set, particularly in light of the historical context of climate variability. Metadata will be included with all DAC quality controlled data sets either as additions to the data or as a part of the QC reports. Ideally, the metadata contains information on the instrument setup, calibrations, and formats for the meteorological data.

Data quality flagging, not value replacement, is the method used to denote suspect or erroneous data. The choice to flag questionable data at the DAC, as opposed to data correction, was made to insure the data remain in the original form that was sent to the DAC. For scientific reasons, the end user must determine appropriate utilization of the flags provided. This will allow additional processing/decisions to be made at later times.

The QC procedure for all WOCE data sets is a four-step process schematically outlined in Fig. 1. The process begins with the acquisition of a WOCE meteorological data set and accompanying metadata (e.g. README file). The first step of the QC is converting the meteorological data to a standard format for internal use at the DAC. The format includes metadata and is described in Section 2. Once converted, the data are sent through an automated preprocessing program that checks for values outside of a realistic range, statistical outliers, unrealistic ship movements, etc.

The preprocessing procedures are outlined in Section 3. During the third step of the QC, the preprocessed meteorological data are visually inspected by a trained Data Quality Evaluator (DQE) using the Visual Data Assessment Tool (VIDAT). VIDAT is an interactive, graphically-based tool developed at the FSU DAC, that allows the DQE to add or remove QC flags from the meteorological data. VIDAT is described in detail in Section 4. When the meteorological data have passed all final inspections, the data are converted to a standard format suitable for public distribution and further analysis. At this time a QC report is written combining flag information from the preprocessor and the DQE's VIDAT evaluation. More information, i.e. calibrations done prior to the data's arrival at the DAC, may be extracted from the metadata and added to the QC report. The resulting quality controlled data and report are then made available for the WOCE community.

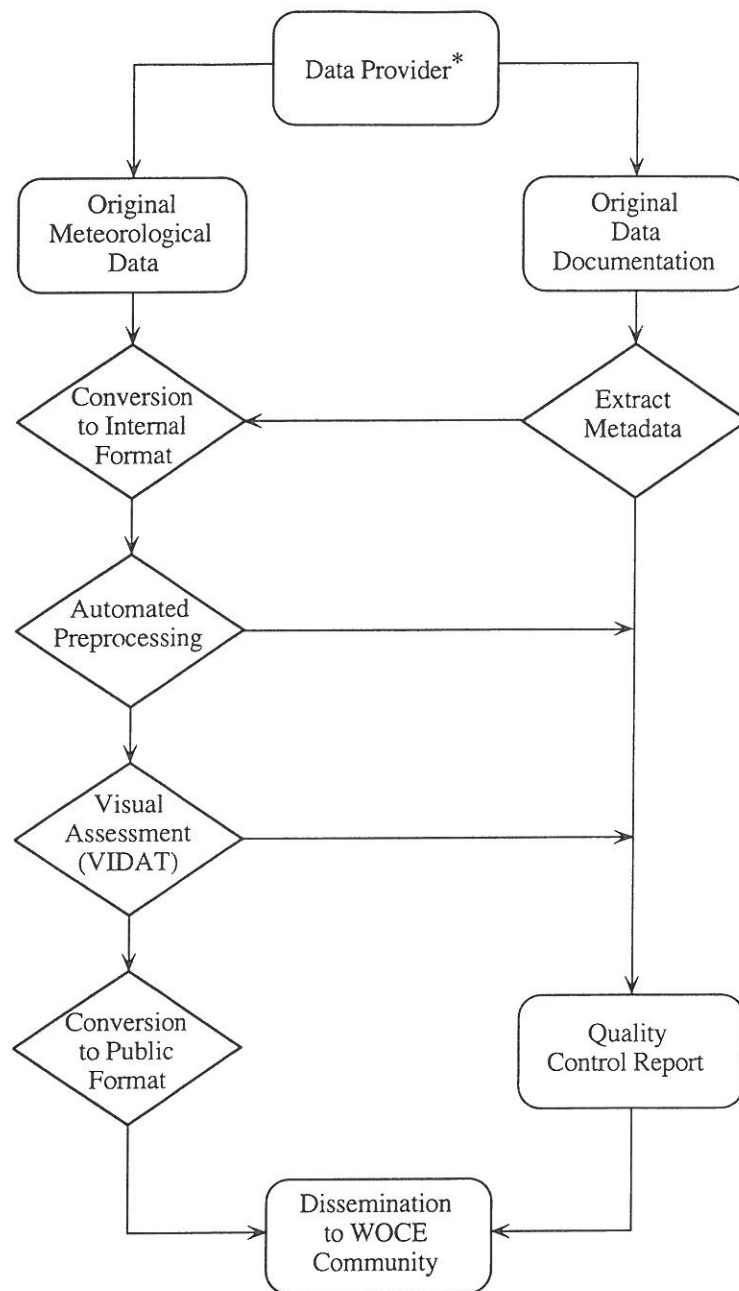


Figure 1. Data flow (arrows) through the FSU DAC. The data that arrive from a provider (*includes individual investigators, chief scientists, data centers, etc.) usually have two components; the data values and the documentation. The data values are converted to a standard format and combined with metadata extracted from the documentation. Automated preprocessing and visual inspection (VIDAT) of the data follow. In the final stages, the data are converted to a user friendly public format and QC reports are written. Both ASCII and netCDF formats of the data and the reports are then distributed to the WOCE community.

2. DATA FORMAT

The DAC for surface meteorology chose to store all data in the network Common Data Format (netCDF) (information regarding netCDF is found in Appendix A). NetCDF was selected to take advantage of its portability across many different platforms, its capability to be self-describing, i.e. both meteorological data and metadata can be stored in a single file, and its growing popularity and acceptance in the atmospheric and oceanic communities. NetCDF is used for all data files internal to the DAC and for the data files available to the scientific community. A file in the American Standard Code for Information Interchange (ASCII) format will also be available to the community.

An explicit description of the internal file format is not presented here (details can be found in Smith and Legler, 1995a). In brief, the netCDF format allows a set of *global* attributes which define information pertaining to the entire netCDF file. The global attributes include a title for the data; the measurement site name, elevation, and alphanumeric identification code; the instrument system used; the facility where the data originated; a global missing value flag; a special value flag; the start and end date of the data; the WOCE Hydrographic Project Office EXPCODE; and a release date indicating when the QC process was completed. Variables in each netCDF file include not only meteorological data (pressure, temperature, etc.), but also a number of supporting variables (observation time, platform position and movement, etc.). Each meteorological variable is accompanied by multiple *variable* attributes. The variable attributes store information about the meteorological data, i.e. units, instrument height, instrument type, etc. A variable containing QC flags, described in more detail below, is also included along with a history variable which logs data flag changes. The history information is primarily for internal use, though pertinent information is transferred to the QC reports. An example of an FSU netCDF file is found in Appendix B.

An additional important global attribute, the Florida State University (FSU) version number, was developed to track the data files as they passed through the various stages of the QC process, Table 1. Intermediate numbers were used at times as some steps in the QC process were completed in multiple stages. Version numbering is automated, i.e. each new file created during the QC process is automatically assigned appropriate version numbers.

Table 1: Version numbers for data files at the FSU DAC. Multiple versions are indicated by a '?'.

| FSU Version | Definition |
|-------------|--|
| 0.0.0 | Data in original format (digital or hard copy) |
| 0.0.1 | Digitized original data (from hard copy) |
| 0.0.9 | Data converted to internal FSU DAC netCDF format |
| 0.2.0 | Preprocessing completed, ready for visual inspection |
| 0.2.? | Visual inspection completed |
| 1.0.0 | First public release to community |
| 1.0.? | Updated public version (metadata added/minor errors corrected) |

The FSU version number is also an integral part of the file names assigned to all data arriving at the DAC. The file naming convention created for FSU WOCE surface meteorological data is based on the ship call sign or buoy World Meteorological Organization (WMO) identification number. This convention takes the form:

ID.YYMMsDnDv###.nc

where the components of the file name include:

- ID: Site identifier containing up to 8 alphanumeric values. Normally a ship's call sign or a WMO number for a buoy. A "." always marks the end of the ID.
- YY: Two digit year of the first record in the file.
- MM: Two digit month of the first record in the file.
- sD: Two digit date of the first record in the file.
- nD: Three digit value indicating number of consecutive days of data in file.
- v### FSU version number (without decimal points).
- .nc Identifier for netCDF files. ASCII files will be identified using .asc.

As an example, the following file name was given to the netCDF version of the PR-14 data from the Chilean research vessel (R/V) *Vidal Gormaz* (call sign CCVG):

CCVG.931107011v100.nc
 <-----> <-----> <-> <-->
 ID 1st # FSU#
 Day Days

The file name reveals that the PR-14 cruise began on 7 November 1993 and ran for 11 days. The version number shows that the data has completed QC and is ready for distribution.

2.1 Data conversion to netCDF

Most of the data that arrive at the FSU Surface Meteorology DAC are converted directly to netCDF using standard FORTRAN or C language calls. Normally the only change made to the data is a conversion to a standard set of units (Table 2) using common conversion tables. The original units are noted and retained in a variable attribute. However, if data arrive with no units attached, and only after every effort to determine their units is exhausted, the data are discarded. Furthermore, any data arriving without an accurate time stamp are discarded only after all attempts to obtain correct time information were exhausted. The data also must have platform position information; thus data lacking latitude and longitude values are discarded only after all attempts to correctly position the data failed. Data collected without a record of time, position, or units are useless to the scientific community. All original data contributions are archived in the event additional information is discovered later that would allow the inclusion of discarded data.

Note that wind direction data are converted to the meteorological convention (Table 2). The meteorological convention is to report the direction *from which* the wind is blowing. This is 180° opposite of the oceanographic convention that gives the direction to which the wind is blowing. Further alterations to the wind data are made when the winds are calm. The WMO convention is used; i.e. calm winds have a direction of zero with a speed of zero and north winds have a direction of 360°. Some data received at the DAC however randomly interchange directions of zero and 360° when the winds are calm, or report a north wind using zero for the direction. All attempts to make the wind direction data consistent with the WMO convention are made and changes are reported in the QC report.

In some cases, additional modifications are made to the original (Version 0.0.1) data. These include ordering the data in a correct time sequence and removing duplicate records whenever possible. As time permits and with adequate metadata, obvious typographical errors can be corrected. A simple example of a typographical error is a single time stamp reporting the year to be 1992 when the remainder of the cruise data are from a line measured in 1990. All modifications made to the original data are carefully documented in the QC report and all original data are retained in the FSU DAC archive.

A final modification to the original data structure may be necessary if there are multiple measurements of the same parameter in one data set. For example, if a ship has 3 different instruments to measure sea temperature (i.e. bucket, intake, and a thermosalinograph) the netCDF file will contain three sea temperature variables (TS, TS2, and TS3) each with their own set of attributes. No modification of the original data is required if the three sea temperatures are all recorded at the same rate; however if the time resolution varies, the time stamp from the sea temperature recorded at the finest temporal resolution will be used as the base time for the netCDF file. For example (Table 3), if the SST is recorded every 5 minutes by

thermosalinograph (TS), every fifteen minutes by intake (TS2), and approximately every hour by bucket (TS3), the second two sea temperature time series have their times matched to the nearest time of the first sea temperature (TS). Severe time mismatches are handled on a case-by-case basis and notes are provided in the QC report.

Table 2: DAC standard units.

| Variable | Units | Convention |
|-------------------|--------------------|---|
| time | minutes | since 01-01-1980 00:00 UTC |
| latitude | degrees | + North, - South |
| longitude | degrees | 0-359 degrees East |
| platform heading | degrees | clockwise from true North |
| platform speed | m s^{-1} | |
| wind direction | degrees | direction wind is blowing from (meteorological convention) |
| wind speed | m s^{-1} | |
| atmos. pressure | mb | |
| air temperature | $^{\circ}$ Celsius | |
| wet-bulb temp. | $^{\circ}$ Celsius | |
| dew point temp. | $^{\circ}$ Celsius | |
| sea temperature | $^{\circ}$ Celsius | |
| relative humidity | percent | |
| specific humidity | g kg^{-1} | |
| precipitation | m m | if a rain rate - mm hr^{-1} |
| radiation | W m^{-2} | |

As a final note on the conversion of incoming data to the FSU DAC netCDF format; the netCDF files are limited to 1440 records, which corresponds to the number of minutes in a day. The 1440 record limit was chosen primarily for convenience and to simplify the operation of the preprocessor and visual inspection software. Thus incoming data files with more than 1440 records were broken into smaller files. Not all files contain a single day's data, but the limit of 1440 records still applies.

Table 3: Example of matching times for multiple measurements of the same variable (e.g. sea temperature) from one platform.

| Time for TS | Time for TS2 | Time for TS3 |
|-------------|--------------|--------------|
| 00:00 | 00:01 | 23:58 |
| 00:05 | | |
| 00:10 | | |
| 00:15 | 00:16 | |
| 00:20 | | |
| 00:25 | | |
| 00:30 | 00:31 | |
| 00:35 | | |
| 00:40 | | |
| 00:45 | 00:46 | |
| 00:50 | | |
| 00:55 | | 00:56 |
| 01:00 | 01:01 | |
| 01:05 | | |

2.2 Flagging philosophy

One of the primary objectives of the FSU DAC for surface meteorology is to provide the community with quality controlled data, i.e. the data would be evaluated and assigned a flag (or flags) which relates information concerning the quality of the data. In the past, data quality information was provided through a series of elaborate and sometimes complicated numerical flags which often were difficult to use and at times ignored. Another problem with some older flagging techniques is that they correspond to entire data records instead of the individual data values. For example, if a station records 20 variables, and only the pressure value is in question, the entire record should not be flagged as erroneous.

At the FSU DAC a very simple approach to QC flags was taken. The quality control flags are single alphabetic characters for each data value in a record that indicate either problems or notable features. The quality control flags for multiple variables in each record at a single time are combined in a single character string and stored in the flag variable for that record. The length of the flag string is equal to the number of variables that underwent quality control. As an example, assume that a FSU DAC netCDF file contains only these five variables: time, latitude, longitude, atmospheric pressure, and air temperature data along with a flag variable. If the

values for the first record of all five variables pass all quality control checks, then the first record of the flag variable will be ZZZZZ (a Z indicates acceptable data - see Appendix C) i.e. one flag for each of the 5 data values. If the second record contains a non-sequential time (flagged by C), but good latitude and longitude values, a pressure value of 1190 mb (too high - flagged by B), and a temperature that is 6 standard deviations from the climatology (flagged by G), then the second record flag string will be CZZBG. The non-Z flags are values that failed one or more of our QC tests (outlined in sections 3 and 4).

In order to gain easy, quick access to flags which correspond to specific variables, each quality controlled variable in a FSU WOCE netCDF file has a variable attribute *qcindex*. The *qcindex* is an integer index that points to the position in the flag string for that variable. In the above example the *qcindex* for pressure is 4, thus the fourth character in the flag string is the QC flag for pressure. The flagging method described here allows access to the flag for any variable using the *qcindex* variable attribute.

2.3 Flags added during data conversion to netCDF

Prior to all QC processing, all acceptable data values in a FSU DAC netCDF file are assigned a Z flag. Three flags may be added during the conversion of the data to the FSU format. The A flag is assigned to any data value where the original units are not known or explicitly provided and the units were determined using supporting information (i.e. neighboring stations, country unit conventions, synoptic maps, etc.). The supporting information used will be outlined in a QC report for the data.

In a few cases, meteorological data arrived at the DAC with separate platform position and movement files. Efforts were made to match the position data accurately to the meteorological data, however, due to varying time stamps a perfect match was not always possible. As a result some of the platform positions and movements are uncertain. In other cases, positions are known to be uncertain according to information from the data provider. The P flag denotes uncertain platform position/movement information.

The third possible flag added during the conversion process, Q, is for data that arrives already flagged as questionable. The DAC attempts to match any quality control flags on the incoming data to a similar flag in our system. When data arrives with flags already assigned, a section is added to the QC report explaining how the original flags were converted to the FSU system.

3. PREPROCESSING

After the incoming data are converted to the standard netCDF format, the first QC procedure applied is an automated preprocessing program. This program (written in standard FORTRAN) was designed to flag data that failed to pass a series of objective evaluations. These tests include, in order of application:

- verifying the existence of time, latitude, and longitude data for every record
- flagging non-sequential and/or duplicate times
- flagging data greater than four standard deviations from a climatology
- flagging data that are not within a realistic range of values
- flagging platform positions and speeds that are unrealistic
- flagging positions where oceanographic platform moves over land
- flagging inaccurate earth relative wind speed and direction data
- flagging data that fail the relationship: air temperature \geq wet-bulb temperature \geq dew point temperature.

The preprocessor, as the name implies, was designed only as a preliminary scan through the data. The final decision to keep, reject, or add any flag falls to the DQE performing the visual inspection (section 4). The primary purpose of the preprocessor is to automate the flagging process and highlight suspect data for the DQE. Normally the DQE will not modify the flags added by the preprocessor, but the DQE does review the flagged data and may modify the flags as needed.

The preprocessor creates two files in addition to the new version of the data file: a diagnostic file and an assessment file. The assessment file contains output from the preprocessor operation, including a tabular list of all flags added to the meteorological data in the netCDF file. This information is reviewed to identify systematic data errors. The diagnostic file stores run-time errors and warnings used to diagnose problems with the preprocessor or the input netCDF file.

During operation, the preprocessor initially inquires about the contents of the netCDF file. The first QC check verifies the existence of the time, latitude, and longitude data. As was mentioned in the data format section, meteorological data that lack time or position are useless to the scientific community. When the preprocessor locates a record with no time, latitude, or longitude, an error message is written to the diagnostic file and the preprocessor halts operation. If no reliable value for the missing time, latitude, or longitude can be found, the entire data record is removed and the preprocessing restarted.

Once the existence of the time, latitude, and longitude are verified, the preprocessor begins screening the individual data values. The preprocessor divides the tests into two groups, univariate checks (involving a single variable) and the multivariate (involving multiple variables) checks, Table 4. All univariate checks are completed before the multivariate checks. In the following sections, the method for each preprocessing check will be outlined.

Table 4: Flag, flag purpose, and flag type issued by preprocessor.

| Flag | Purpose | Type |
|------|--|--------------|
| B | Value out of realistic range | Univariate |
| C | Time not sequential | Univariate |
| D | Failed $T \geq T_w \geq T_d$ test | Multivariate |
| E | Failed resultant wind recomputation test | Multivariate |
| F | Platform velocity unrealistic | Multivariate |
| G | Value > 4 standard deviations from climatology | Univariate |
| L | Oceanographic platform crosses land | Multivariate |
| T | Time duplicate | Univariate |

3.1 Time sequential/duplicate tests

The first of the univariate tests flags any observation where the time is not sequential or is duplicated. The time is flagged with a *C* when the next record does not have a time later than the current record. Often when multiple out of sequence times occur in a file, efforts are made to reorder the records. If the original data order is modified, a note will be included in the QC report.

When two consecutive the times are identical, the time duplicate flag *T* is added to both times. The time duplicate check only indicates that the times are identical. Visual inspection of the remainder of the data record is used to determine if they are exactly identical records. If both entire records are identical, one is discarded. However if any discrepancies occur in the two suspect records with duplicate times, both are retained, and assigned *T* flags.

An obvious limitation of the time tests is that the first time value is assumed to be correct. The major advantage to the time tests is that they allow the correction, in most cases, of the time sequence and the removal of duplicate records.

3.2 Statistical test

The second univariate test compares the data values to a representative climatology. The climatology used was created by da Silva et al. (1994) and is on a 1° by 1° grid over the global oceans. This climatology includes both mean ($\overline{X_c}$) and standard deviation (s.d.) data based on the Comprehensive Ocean-Atmosphere Data Set (COADS) and was adjusted for wind speed bias and weather observation codes.

The statistical test is applied as follows: 1) the data value (*x*) to be compared to the climatology is mapped to the nearest climatology box using the data value's

latitude and longitude, 2) if x falls outside the range defined by the $(\overline{X}_c) \pm 4 \text{ s.d.}$, x is flagged with a G. The statistical test was designed primarily to focus the DQE's attention on extreme values. The DQE may change the G upon visual inspection, e.g. when a physical explanation for the extreme values can be determined.

For example, atmospheric pressure from the island station Nauru, clearly has values less than 4 s.d. (dashed line) from the climatological mean (solid line) Fig. 2. All values below the dashed line were flagged with a G. In the example shown, no reason for the low pressures could be determined, so the G flags were unchanged to caution other users of the Nauru pressure data.

The main limitation to the statistical test is the test can only be applied to variables that are in both the climatology and the netCDF file. In the case of the da Silva et al. (1994) climatology, the statistical check is limited to the wind speed, sea-level pressure, air temperature, sea surface temperature, and relative humidity data. Another limitation of the comparison occurs due to varying observational systems. For example, the sea temperature can be measured from instruments located anywhere from the surface to approximately 4 m below sea-level. These measured sea temperatures are compared to a climatological SST based primarily on bucket and intake temperatures. The same instrument related errors can apply to all the variables since ship instruments are often not at standard heights or in standard shelters.

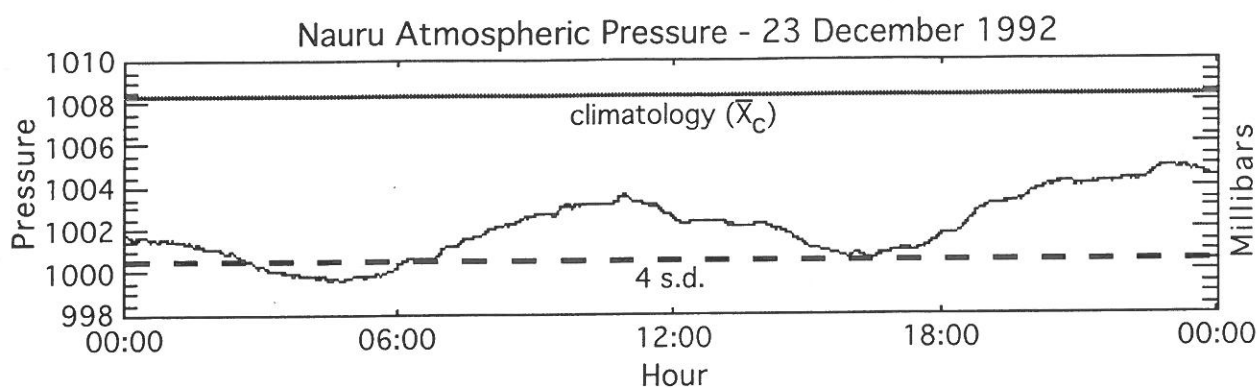


Figure 2. Atmospheric pressure data from the island of Nauru for December 1992 showing values that are greater than four standard deviations (dashed line) from the da Silva et al. (1994) mean (grey line).

3.3 Bounds test

The final univariate test in the preprocessing determines whether the data values fall within a physically realistic range. Acceptable ranges vary depending on the meteorological variable, Table 5. The flagging strategy is simple, if a value falls outside the acceptable range, a *B* flag is assigned to the value. The ranges chosen by the DAC were designed to indicate extreme values. Some of these out of bounds values, for example an air temperature of -15.0°C near the Antarctic Coast, are realistic and the bounds flag is removed by the DQE.

Table 5: Range bounds used by DAC.

| Variable | Lower Bound | Upper Bound | Units | Comments |
|-----------------------|-------------|-------------|--------------------------|------------------|
| time | 1-1-1980 | 12-31-1999 | | |
| latitude | -90 | 90 | degrees | |
| longitude | 0 | 359 | degrees | |
| platform heading | 0 | 359 | degrees | |
| platform speed | 0 | 15 | m s^{-1} | research vessels |
| | 0 | 0 | m s^{-1} | stationary buoys |
| | 0 | 2 | m s^{-1} | drifters |
| plat. wind direction | 0 | 360 | degrees | |
| plat. wind speed | 0 | 40 | m s^{-1} | |
| wind direction | 0 | 360 | degrees | |
| wind speed | 0 | 40 | m s^{-1} | |
| pressure | 950 | 1050 | mb | Sea-level |
| air temperature | -10 | 40 | $^{\circ}\text{Celsius}$ | |
| wet bulb temperature | -10 | 40 | $^{\circ}\text{Celsius}$ | |
| dew point temperature | -10 | 40 | $^{\circ}\text{Celsius}$ | |
| sea temperature | 0 | 35 | $^{\circ}\text{Celsius}$ | |
| relative humidity | 0 | 100 | percent | |
| specific humidity | 0 | 48 | g kg^{-1} | |
| rain rate | 0 | 150 | mm hr^{-1} | |
| radiation | 0 | 1400 | W m^{-2} | |

For example, some relative humidity data collected at Manus Island exceed the normal maximum of 100%, Fig. 3. Since 100% is the upper bound for relative humidity (Table 5), *B* flags are assigned to these range points. The values in Fig. 3 only slightly exceed the upper bound for relative humidity which implies that the sensor could be either slightly off calibration or possibly located in a region with dense fog.

The bounds test also verifies that coded data are within expected ranges. For example, cloud type codes should be in a range of 1 to 10. If a value is not in that range, the bounds routine changes the out of range data value to the special value (-8888) and notes the change to the assessment file. The assessment file is then checked by the DQE to search for systematic errors in the codes. The change in the data is made to eliminate erroneous codes that are of no use to the end user, and the check is done only for coded variables that *do not* have quality control flags assigned to them.

The bounds test highlights extreme events in all variables, not just the five variables covered by the statistical test. The bounds test has been quite reliable in noting realistic extreme events like typhoons and arctic outbreaks. One disadvantage is that the routine currently uses the same bounds over the entire globe. A better scheme would vary the range of values for different climatic regions of the globe.

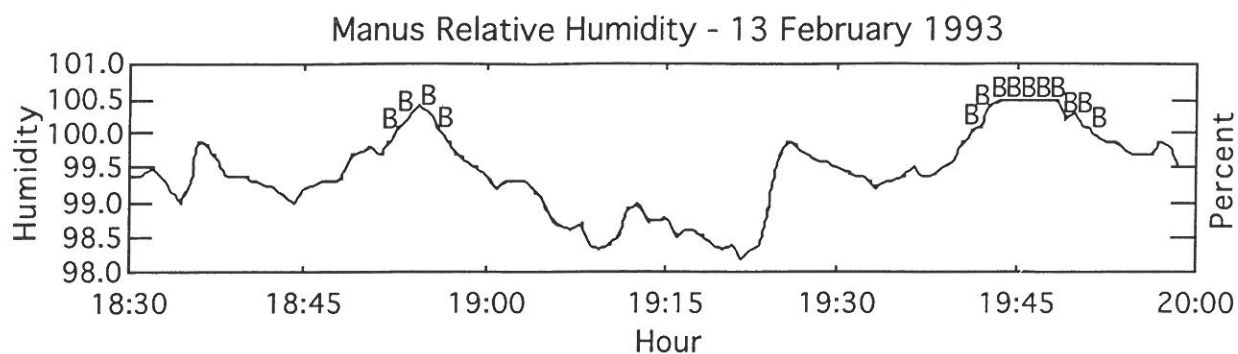


Figure 3. Relative humidity data from Manus island for February 1993. The *B* flags indicate values that exceed the upper bound for relative humidity (100%).

3.4 Platform speed test

The first of the multivariate tests performed by the preprocessor determines if the speed of the platform exceeds a selected realistic threshold velocity (Table 6) for the platform. Platform speed is determined using the latitude and longitude positions of the platform since all WOCE DAC netCDF files will have latitude and longitude data, but not all data provided to us includes platform speed and heading information. If the speed exceeds the threshold, both the latitude and longitude are flagged with an F . All F flags are verified by the DQE when the platform speed and heading are present.

Table 6: Threshold speeds used to assign unrealistic platform speed flag (F).

| Platform Type | Maximum Speed (m s ⁻¹) |
|-----------------|------------------------------------|
| Research Vessel | 15.0 |
| Drifting Buoy | 2.0 |
| Anchored Buoy | 0.0 |

The platform speed is determined using the great circle calculation on the spherical surface of the earth. Given two positions of a platform (ϕ_1, θ_1) and (ϕ_2, θ_2) , where ϕ is the latitude and θ is the longitude, then the radii from the center of the earth to these two surface points are

$$\vec{r}_1 = r \cos \phi_1 \cos \theta_1 \hat{i} + r \cos \phi_1 \sin \theta_1 \hat{j} + r \sin \phi_1 \hat{k} \quad (1a)$$

$$\vec{r}_2 = r \cos \phi_2 \cos \theta_2 \hat{i} + r \cos \phi_2 \sin \theta_2 \hat{j} + r \sin \phi_2 \hat{k} \quad (1b)$$

The dot product of the two radii can be found using

$$\vec{r}_1 \cdot \vec{r}_2 = |\vec{r}_1| |\vec{r}_2| \cos \gamma = r^2 \cos \gamma \quad (2)$$

where γ represents the angle between the two vectors in radians. Another definition of the dot product takes the sum of the products of the coefficients for each unit vector,

$$\vec{r}_1 \cdot \vec{r}_2 = r^2 \cos \phi_1 \cos \theta_1 \cos \phi_2 \cos \theta_2 + r^2 \cos \phi_1 \sin \theta_1 \cos \phi_2 \sin \theta_2 + r^2 \sin \phi_1 \sin \phi_2 \quad (3)$$

Equating (2) and (3) gives

$$r^2 \cos \gamma = r^2 \cos \phi_1 \cos \theta_1 \cos \phi_2 \cos \theta_2 + r^2 \cos \phi_1 \sin \theta_1 \cos \phi_2 \sin \theta_2 + r^2 \sin \phi_1 \sin \phi_2 \quad (4)$$

Dividing (4) by r^2 and taking the arccosine results in

$$\gamma = \arccos \left\{ \cos \phi_1 \cos \theta_1 \cos \phi_2 \cos \theta_2 + \cos \phi_1 \sin \theta_1 \cos \phi_2 \sin \theta_2 + \sin \phi_1 \sin \phi_2 \right\} \quad (5)$$

The distance, d , between the two positions on the globe is then found by

$$d = r\gamma \quad (6)$$

The speed is then found by dividing the distance by the number of minutes elapsed between the two platform positions.

For all data with a time step of at least three minutes, velocities are calculated using sequential latitude and longitude values. Empirical tests showed that three minutes was the smallest useable increment for the latitude and longitude data. When the time step is one minute, too many positions are flagged because the distances traveled by the platform are so small that dividing the distance by the time often results in speeds nearing infinity. Thus one minute data were tested using three minute intervals.

For example, latitude and longitude plots for the *R/V Knorr*, Fig. 4 indicate a region of unrealistic ship speeds. The data show rapid one minute changes in latitude from 53°S to 50.4°S near 1200 UTC on 23 February 1993. Since the data on the *R/V Knorr* are recorded every minute, the 3 minute average speed centered near the sudden latitude jump would be nearly 1584 m s⁻¹. This far exceeds the maximum speed for a research vessel, Table 6 . As a result, the preprocessor flagged both latitude and longitude positions used to calculate the speed. Again the preprocessor highlights a suspect portion of the data and the DQE will make the final determination on the data values to flag using the visual editor.

A major limitation of the current speed test is that the accuracy of the calculations is limited by the accuracy of the latitude and longitude data. Normally latitude and longitude are stored to the hundredth of a degree which is equivalent to 1100 m. To determine a ship speed within a threshold of 15 m s⁻¹, more than one minute spacing between sequential positions is needed; thus the 3 minute intervals mentioned previously.

3.5 Land test

The second multivariate test performed on the platform position data verifies that an oceanographic platform does not move over land. As an example, assume a fisherman gets a drifting buoy tangled in his nets. Instead of trying to untangle the mess at sea, the fisherman hauls the net and buoy onto his boat and returns to shore. If the buoy continues transmitting, the position will now be reported as over land.

The more common occurrence, however, is simply erroneous latitude and longitude data. For example, latitude and longitude positions for the *R/V Kexue1* on 1 January 1993, Fig. 5 indicates many of the ship positions are over Australia. The most plausible explanation for the erroneous positions is that the latitude positions

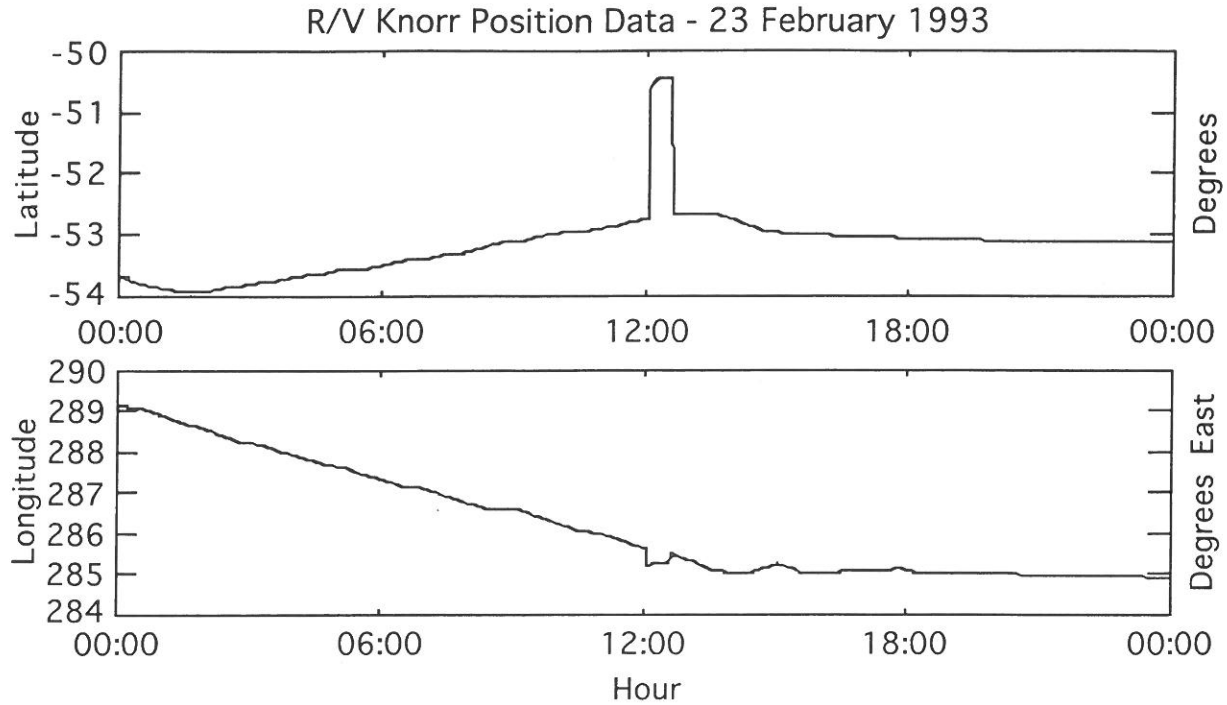


Figure 4. Rapid changes in latitude and longitude values, like these for the R/V *Knorr*, are flagged for failing the platform speed test. All values within a region of rapid change are flagged.

should be in the Northern Hemisphere, not the southern. Incorrect negative signs are a common problem with position data.

The test for movement over land is similar to the statistical test. The first step was to create a land/ocean map of the entire globe. Using the El Topo5 five minute global topography (NOAA-NGDC) we created a land mask and assigned all grid boxes with elevations greater than or equal to zero meters to a value of 1. All negative elevations were set to 0. The result was a 5 minute binary grid. The land test maps the platform's latitude and longitude positions to the appropriate 5 minute box in the land mask; and the land mask is checked for indications of land. When the position occurs over land, the latitude and longitude values are both flagged with an "L".

The 5 minute (~9 km) resolution of the land mask is the main limitation to the land test. When a research vessel enters port or passes through narrow channels (i.e. Panama Canal, Straights of Magellan, etc.), land flags are often assigned by the preprocessor. The DQE always verifies the land flags and removes any that are the result of passage through small bodies of water.

The land test often points out possible sign errors in the latitude and longitude data. The erroneous position data can then be modified after verifying the actual cruise position (via email or other contacts). No changes to position data due to probable sign errors have currently been applied, however any changes to the position data will be documented in the QC reports.

Positions of the R/V Kexue1 – 1 January 1993

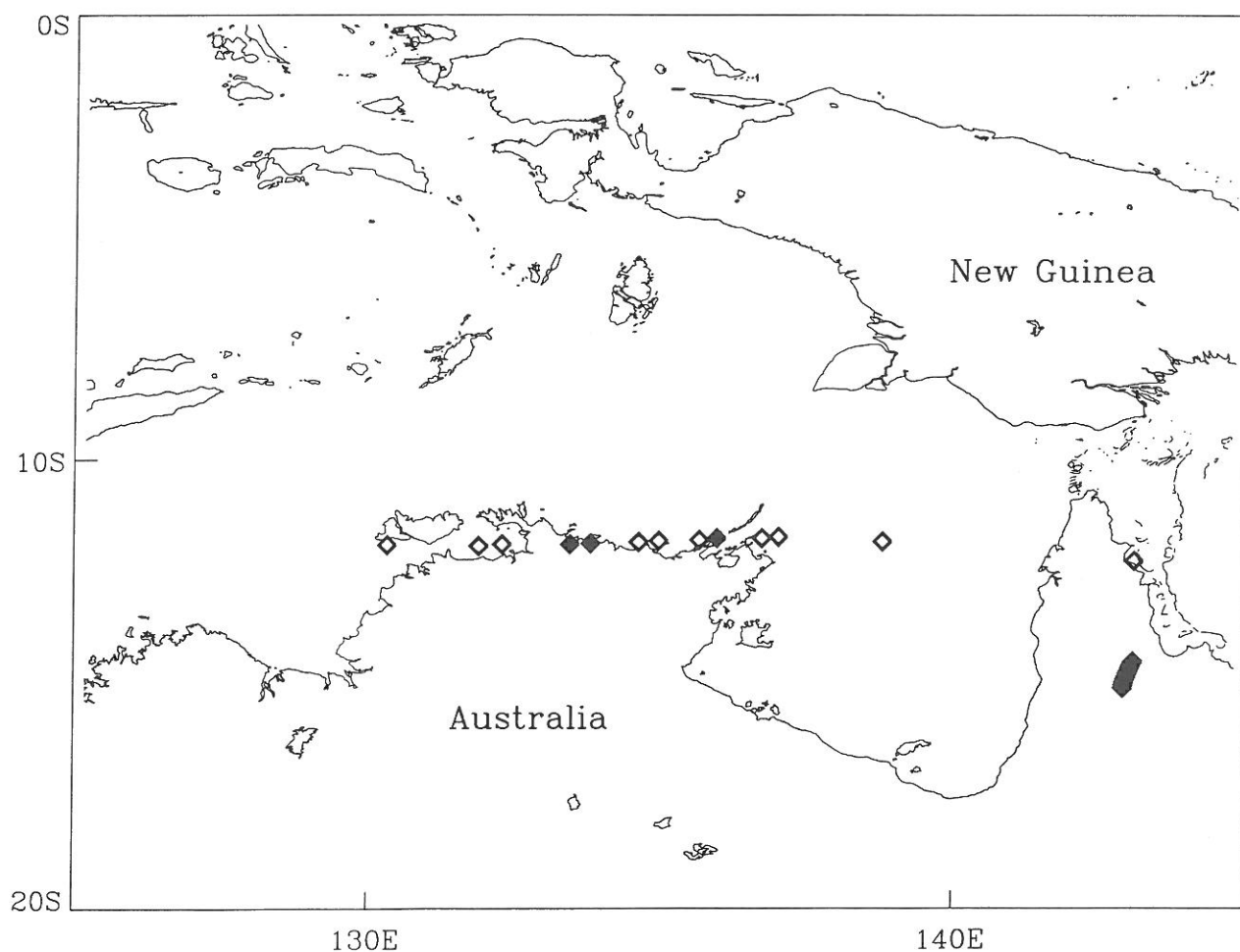


Figure 5. Position data for the R/V *Kexue1* which resulted in land flags at all the positions over Australia (filled diamonds). The open diamonds are positions the land test found to be over water

3.6 Earth relative wind recomputation test

The third multivariate test verifies the values of the reported earth relative winds. The test is accomplished by recalculating the earth relative winds, hereafter referred to as 'true' wind, and comparing the calculated value to reported true winds. Before outlining our test parameters a discussion of the derivation of the true wind is needed. Simply stated, the true wind (T) is derived by subtracting the wind induced by the motion of the platform ($-C$; as the motion of buoys is negligible, hereafter we refer to ships) from the ship relative wind vector (S ; hereafter referred to as the 'ship' wind). This calculation can be achieved using a graphical vector subtraction technique (Fig. 6a) or a numeric approach.

Care must be taken when using either approach as the orientation of the ship with respect to true north must be taken into account. The orientation of the ship is normally represented by the heading of the vessel, which in most cases does not equal the course of the vessel over the ground. At low ship speeds, currents and the wind will push a ship sideways through the water. As a result, the vector of the ships motion over the fixed earth rarely lies in the direction of the bow of the vessel.

A second key point regards ship winds, as it is critical to know a reference point from which the winds are measured. One common practice is to make the bow zero degrees. Other vessels however, may orient the winds with the stern being zero. Thus all vessels should report the zero reference line for their ship winds. It is worth noting that this may or may not be the direction of the zero line on the anemometer itself (Leslie Hartten, personal communication, 1996).

As a result, the WOCE DAC has determined that six parameters are necessary to accurately calculate the true wind from any vessel. These parameters include the ship wind direction and speed, the course and speed of the vessel, the heading of the vessel, and the zero reference for the ship wind direction. If provided, these six variables can be used in the following numerical calculation of the true wind.

Prior to discussing the true wind equations it is necessary to define the two coordinate systems that are used. The first is the earth coordinate system which has zero degrees on the positive y-axis and degree values increasing in a clockwise direction. The second is the cartesian coordinate system that has zero degrees defined on the positive x-axis with degree values increasing in the counter-clockwise direction. For this discussion, lower case variables are used to denote the earth coordinates and upper case variables for cartesian coordinates.

Figure 6b is a schematic representation of a research vessel in the earth coordinate system. The vessel has a heading (h) of 45.0° , a course (c) of 30.0° , and speed over the ground (ss) of 5.0 m s^{-1} . The ship wind speed (ws) is 10.0 m s^{-1} and is blowing to a direction (d) of 145.0° . The ship wind direction is referenced to a zero degree line (z) pointing to starboard on the vessel ($z = 90.0^\circ$ with respect to the bow).

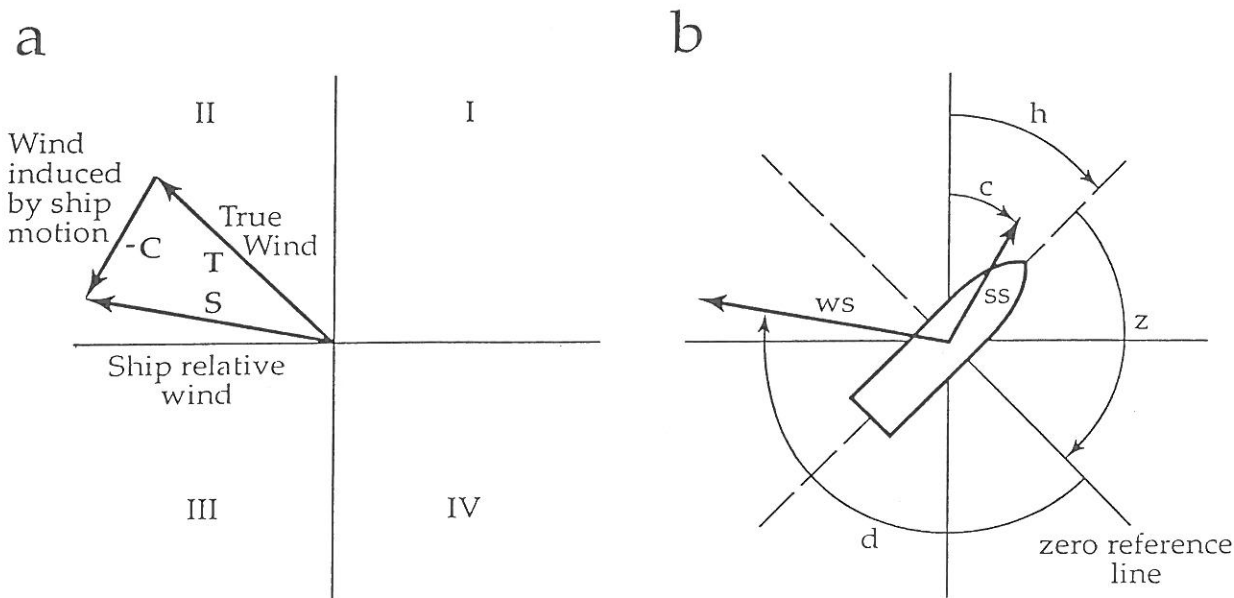


Figure 6. Schematic diagrams showing a) vector subtraction method of calculating the true wind and b) the six vectors and angles required to compute the true wind (T). In a) the roman numerals mark the quadrants in the cartesian coordinate system. (see text for more details)

To begin the calculation, the ship wind is referenced to true north in the earth coordinate system. Referencing the ship wind is accomplished by summing the heading, zero line angle, and ship wind direction. The resulting angle is then converted to cartesian coordinates using

$$S_{\theta} = 90^{\circ} - (h + z + d) \quad (7)$$

where S_{θ} is the ship wind direction in cartesian coordinates (i.e. referenced to true north). Next the earth relative course (c) of the ship must be expressed in cartesian coordinates using

$$C_{\theta} = 90^{\circ} - c \quad (8)$$

where C_{θ} is the course angle in cartesian coordinates. The true wind is then computed by summing the components of the ship relative wind and course

$$T_u = ws \cos S_{\theta} + ss \cos C_{\theta} \quad (9a)$$

$$T_v = ws \sin S_{\theta} + ss \sin C_{\theta} \quad (9b)$$

where T_u and T_v are the north-south and east-west components of the true wind. The true wind speed (tspd) and direction (tdir) in the earth coordinate system can then be calculated from 9a and 9b, with the true wind speed being

$$\text{tspd} = \sqrt{(T_u^2 + T_v^2)} \quad (10)$$

and the true wind direction being

$$\text{tdir} = 270^\circ - \text{atan} (T_v / T_u) . \quad (11)$$

The 270° in (11) converts the value of $\text{atan} (T_v / T_u)$ to a direction *from* which the wind is blowing (meteorological convention) in the earth coordinate system. Care must be taken in this calculation due to the nature of the arctangent function. If (11) is computed using the FORTRAN function 'atan2 (T_v , T_u)', the correct value of the true wind direction will be returned. However, if the arctangent function "atan" is used, 180° must be subtracted from the value of (11) when the true wind vector falls in quadrant two or three in the cartesian coordinate system (quadrants are noted as roman numerals in Fig. 6a).

Returning to the example outlined in Fig. 6b, the conversion to cartesian coordinates using (7) and (8) results in values of -190° and 60° for S_θ and C_θ respectively. Computing the wind components using (9a) and (9b) gives an $T_u = -7.3 \text{ m s}^{-1}$ and an $T_v = 6.1 \text{ m s}^{-1}$. Using (10), the earth relative wind speed is 9.5 m s^{-1} . Because T_u is negative in this example, the angle computed from (11) will be in the second cartesian quadrant (Fig. 6a) and 180° must be subtracted to the result of (11). The resulting true wind direction is blowing *from* 129.9° .

For WOCE quality control, the above calculation is used to compute true wind values received at the DAC. After the calculation is complete, a comparison between the true winds reported by the research vessel and the DAC computed winds is made. If the directions differ by more than 10° , the true wind direction reported by the vessel is flagged with an E. For wind speed, an E flag is applied when differences of more than 5 m s^{-1} occur. In general, wind direction is flagged more often than wind speed due to a larger variability in the wind direction.

Overall, this test is rarely applied. The primary reason is that very few research vessels record and report the six necessary calculation parameters. The more advanced automated instrument systems, i.e. IMET, multimet, etc., are designed to measure all the parameters necessary for determining the true wind. However, instrument failures or insufficient data parameters often limit the ability to calculate the earth relative wind for comparison to the reported true wind.

It should be noted that whenever a vessel reports only *ship* winds to the DAC as well as the other 4 necessary values, the DAC computes true winds using the method described above and places the true wind values in the WOCE data files.

3.7 $T \geq T_w \geq T_d$ test

The final multivariate preprocess tests the physical principal that the air temperature is always greater than or equal to the wet-bulb temperature which in turn is always greater than or equal to the dew point temperature ($T \geq T_w \geq T_d$; Wallace and Hobbs, 1977). The preprocessor tests the three values as pairs in the following order, $T \geq T_w$, $T \geq T_d$, and $T_w \geq T_d$. If any two of the temperatures exist in the WOCE netCDF file, the appropriate pair test will be accomplished. When all three temperatures exist, all three pair tests are done. Failure of a pair test results in both temperatures in the test being flagged with a *D*. The flag must be added to both temperatures in a pair because the offending temperature cannot usually be determined by a simple logical test. The DQE verifies all the *D* flags.

An example time series, Fig. 7, shows data flagged by the $T \geq T_w \geq T_d$ test for the SR-01 cruise of the R/V *Vidal Gormaz*. At both times when flags were assigned, the data values for air temperature and wet-bulb temperature failed the $T \geq T_w$ test. In the first case the $T=1.5^\circ\text{C}$ and the $T_w=2.5^\circ\text{C}$ and for the second case $T=6.0^\circ\text{C}$ and $T_w=7.5^\circ\text{C}$. These flags were reviewed by the DQE and since no supporting information for typographical or other errors was available, the flags were retained.

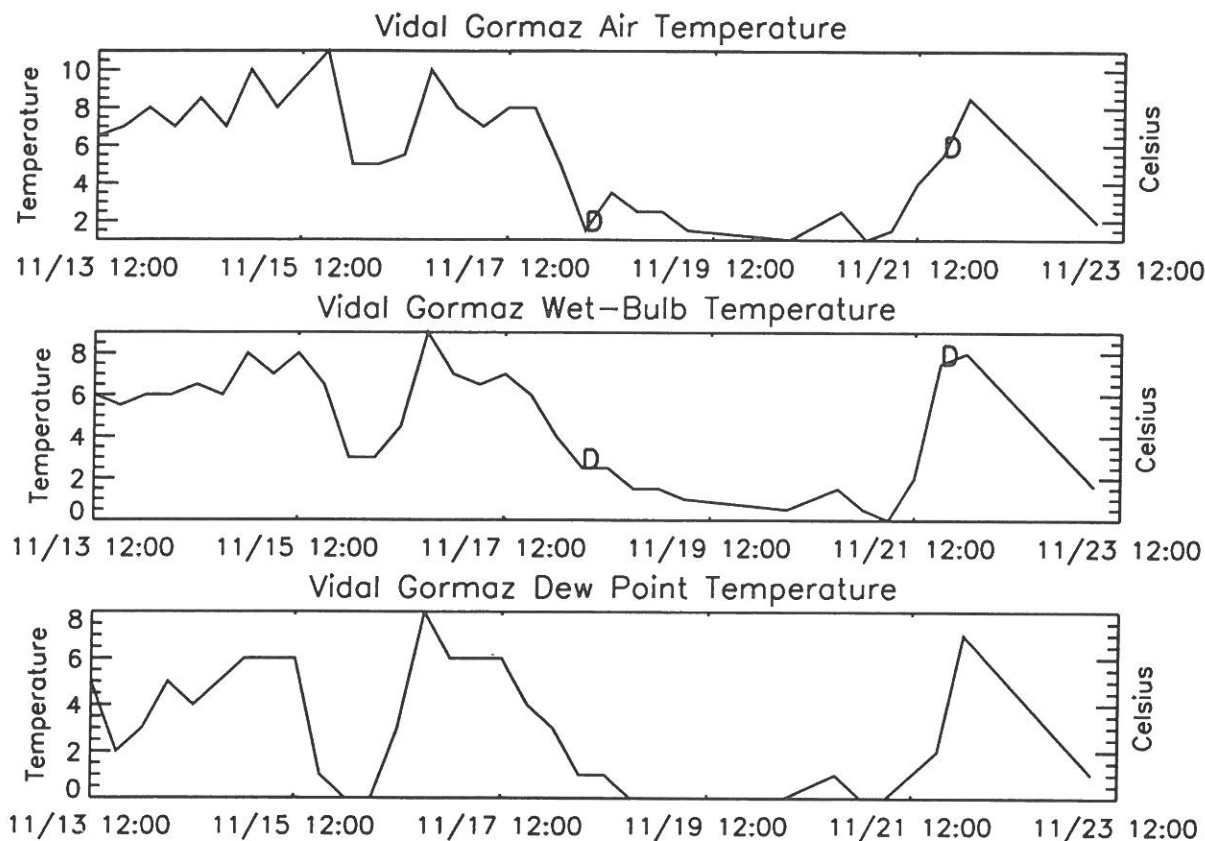


Figure 7. Dry air, wet-bulb, and dew point temperature data from the R/V *Vidal Gormaz*. The values flagged with a *D* failed the $T \geq T_w \geq T_d$ test (see text).

The advantages of the $T \geq T_w \geq T_d$ test are that it highlights physically unrealistic data and can identify typographical errors. The major disadvantage is that the DQE must still review and verify all D flags.

3.8 Order of precedence for preprocessing flags

The preprocessor assigns up to eight different flags (Table 4) to a WOCE netCDF surface meteorology file. Since any data value can only be assigned one alphabetic flag, an order of precedence was created for the flags. In general, multivariate flags will overwrite univariate flags.

More specifically, the time variable may be assigned two different univariate preprocessing flags, time not sequential (C) and time duplicate (T). The methodology of the time test gives the time duplicates flag precedence over non-sequential time flag, i.e. the duplicate test is executed after the sequential test, thus any C flags are overwritten with a T when a duplicate is found. If the duplicates are exact, one is removed, and a second run of the preprocessor will then identify the non-sequential times. Of the two univariate flags not related to time, the bounds (B) flag takes precedence over the statistical (G) flag. This order arose primarily because more variable types can be checked by the bounds test than the statistical test.

In the multivariate tests, only the platform speed and land tests apply to the same variables (latitude and longitude) with the land flag (L) having priority over the platform velocity (F). A platform moving at an unrealistic speed is irrelevant if the platform location is over dry land.

4. VISUAL QUALITY CONTROL PROCEDURES

Once the preprocessing is complete, the WOCE netCDF file is passed to the DQE for visual inspection. The DQE is responsible for reviewing all flags assigned in the preprocessing. Additionally, all data are examined visually to insure the data are of uniform quality. The DQE is a meteorologist with a background in synoptic meteorology, a skill in data assessment, and who has been given a series of training exercises prior to any WOCE data evaluation. The DQE utilizes VIDAT (VIsual Data Assessment Tool), an interactive graphical data display and evaluation tool developed using the Interactive Data Language (IDL; commercially available from Research Systems, Inc.) and written by Jiraporn Whalley. The primary design criterion for VIDAT was to provide the DQE a simple yet extensible interface to flag data displayed in time series format.

A graphical user interface is common throughout VIDAT, and file access is achieved using a point and click file browser. The DQE selects the directory and file (up to 5 files at a time) to be edited. Having multiple files displayed allows the DQE to intercompare data from sequential dates or nearby locations. Once a file is opened, VIDAT lists the variables available within that file for the DQE to select for viewing and editing (Fig. 8).

VIDAT was also designed to record all changes made to the file, create automated reports, and update file versions whenever an edited file is saved. All changes made to the opened files are stored in memory until the DQE saves the new version of the netCDF file. Storing changes in memory allows the DQE to modify the same flag more than once if desired without modifying the original file. When saved, VIDAT combines the input file with the changes stored in memory, thus creating an updated file with an incremental version number.

Three different windows for viewing data are presented: the map plot, the multi-file plot, and the editor. Each offers the DQE new insight into the data undergoing QC. By synthesizing the information provided in the three windows, the DQE can make an informed decision about the quality of any value in the data set.

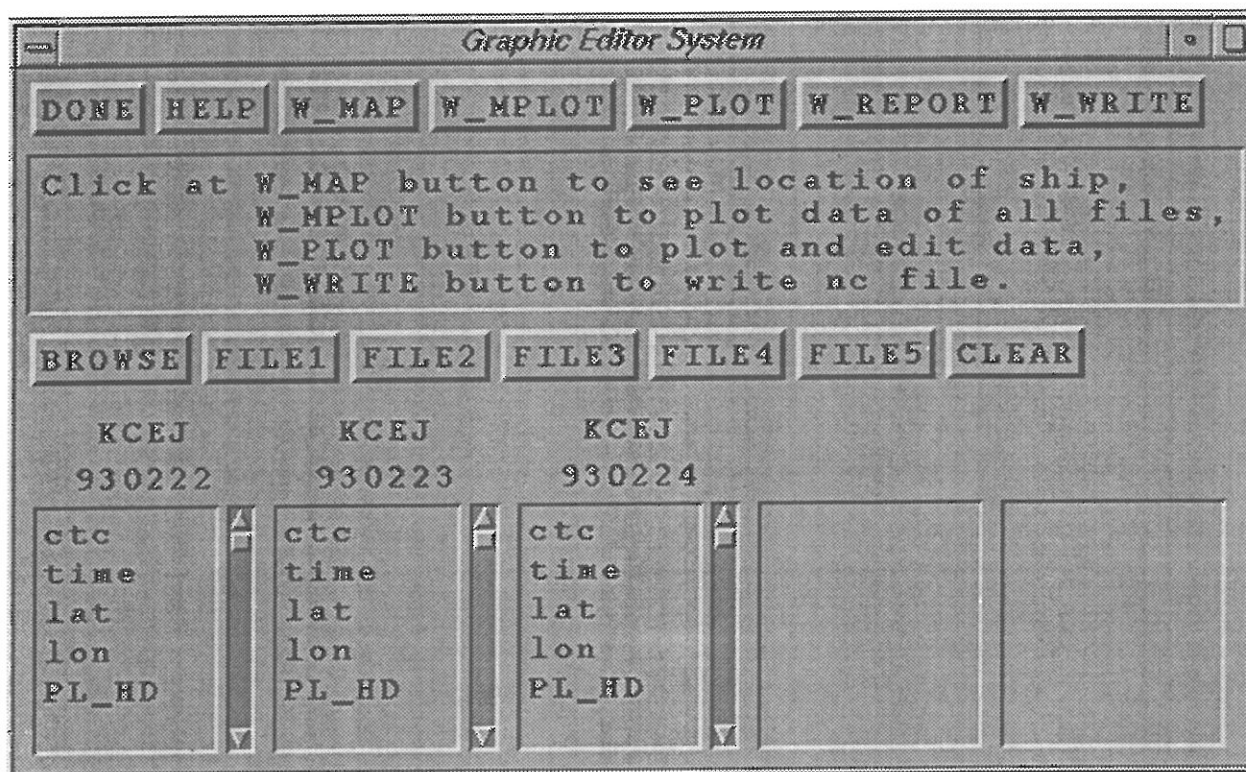


Figure 8. Screen shot of the file access window for the Visual Data Assessment Tool (VIDAT). The buttons across the top have drop down menu options. Three consecutive daily files from the R/V *Knorr* have been selected for review.

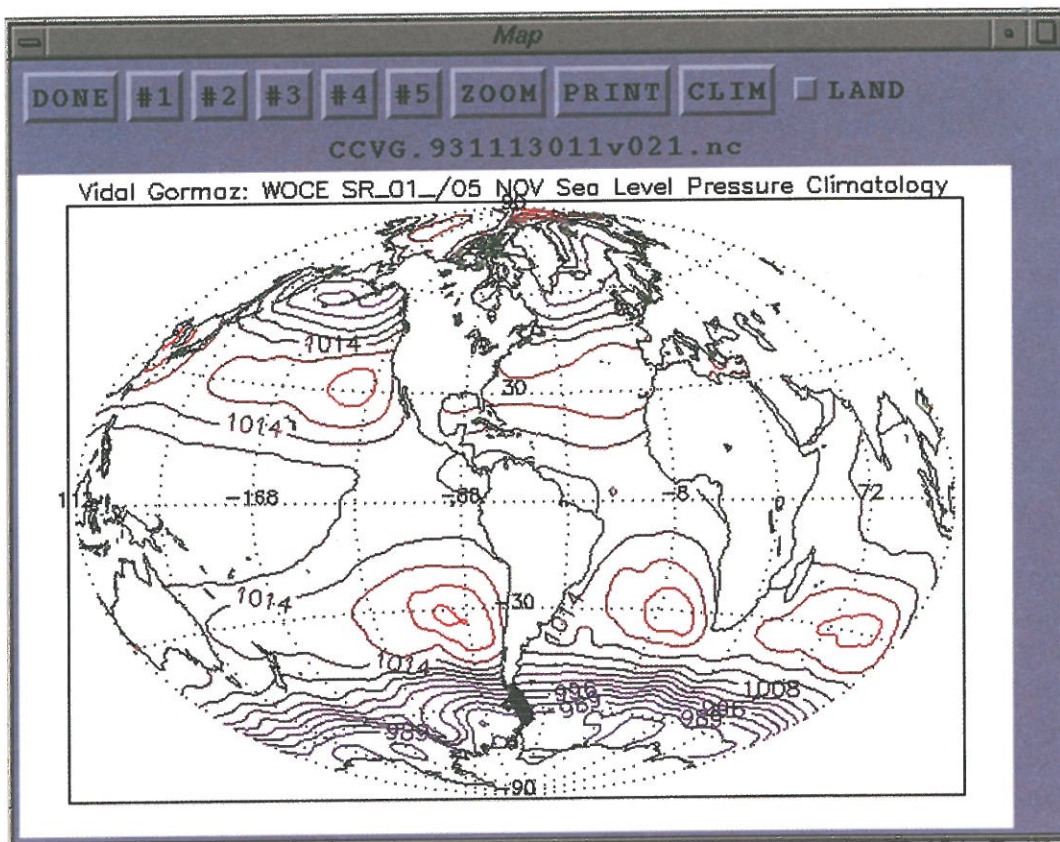
4.1 Map window

The map plot window was designed to map the position of the platform in time on the global ocean. Positions are mapped using the latitude and longitude data from the platform. There is an option to zoom in on a portion of the global map to obtain more detailed information on the platform's movement.

Another feature of the map plot window is the overlay display of the da Silva et al. (1994) climatology over the oceans. Climatologies for wind speed, air temperature, sea surface temperature, precipitation, SLP, relative humidity, and total cloud amount are available. Plotting the climatology under the platform's motion track offers the DQE guidance for flagging data in regions of the globe that may be unfamiliar.

A global map plot of the cruise track for the R/V *Vidal Gormaz* during the 1993 WOCE SR-01 line, Fig. 9a, shows the cruise track clearly marked between South America and the Antarctic Peninsula. The SLP climatology is mapped over the oceans and shows the R/V *Vidal Gormaz* crossing the sharp pressure gradient associated with the circumpolar trough. A zoomed view of the cruise track, Fig. 9b, with numbers (1-4) identifies the direction the vessel moved along the cruise track. Note also that the zoomed cruise track clearly displays an erroneous position along the southbound leg of the cruise.

a)



b)

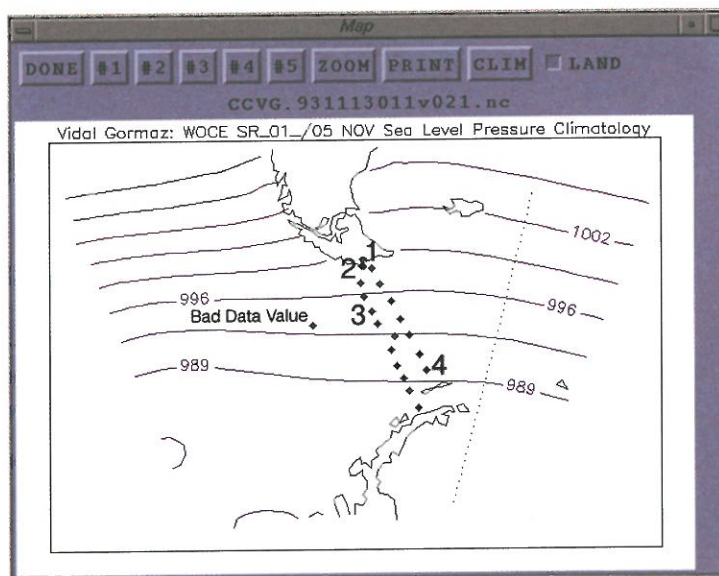


Figure 9. VIDAT screen shot of (a) the full globe cruise track map displaying the 1993 SR-01 cruise of the R/V *Vidal Gormaz* and (b) a zoom of the same cruise track showing an erroneous ship position value. Both panels display the atmospheric pressure climatology for November (da Silva et al. 1994).

4.2 Multiple file plot

It is often necessary to look at time series of one minute data over several days which, due to the file size limitation of 1440 records, can extend over several files. The multiple file plot allows the DQE to select a variable and plot the time series of that variable using the data from all open files. A multi-plot example of temperature data for five days from the R/V *Knorr*, 00:00 UTC on 21 February 1994 through 00:00 UTC on 26 February 1994, shows the trends in the temperature data on time scales greater than one day (Fig. 10).

The multiple plot window also allows an initial investigation of entire data records from a specific platform, identifying major problems, before beginning the review/QC process. Furthermore, the climatology time series corresponding to the appropriate month and location for wind speed, air temperature (Fig. 10 - dotted line), sea surface temperature, precipitation, SLP, relative humidity, and total cloud amount can be plotted over the data time series. The climatology assists in determining whether the data values are representative for the location of the platform.

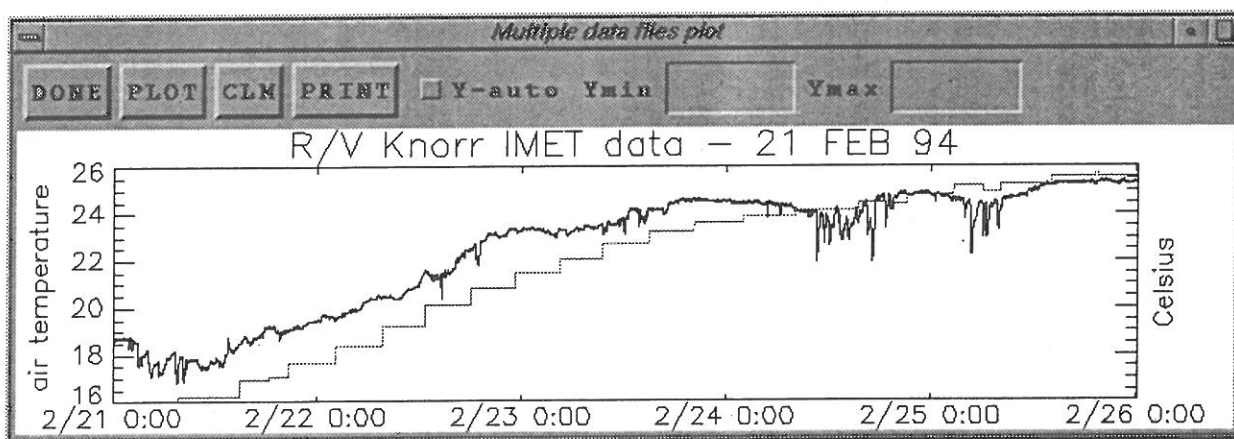


Figure 10. Screen shot of the VIDAT multiple file display window. The window shows multiple files (days) of one minute air temperature data collected by the IMET system on the R/V *Knorr*. The dotted line represents the February air temperature climatology along the same cruise line.

4.3 Editor

The VIDAT editor allows the DQE to modify and add flags to FSU netCDF files. The editor was designed as a point and click interface for highlighting and flagging data values for multiple variables from up to five files, Fig. 11. This example shows the time series plots for wind speed, air temperature, and sea temperature recorded by the R/V *Vidal Gormaz* from 13 to 23 November 1993. Up to six selected variables may be viewed at one time. The variables available for editing appear on the buttons to the right of the plot button. The editing window only allows three of the six plots to be viewed at one time, although the other three variables can be viewed by scrolling.

Other features of the editor include zooming on specific time windows, y-axis modification, and overlaying climatology. Similar to the multiple plot window, climatology time series for wind speed, air temperature, sea surface temperature, precipitation, SLP, relative humidity, and total cloud amount can be displayed automatically for the correct month, location, and variable on any of the six graphs.

The time series plots in the editor can be viewed either with (Fig. 11) or without (Fig. 12) flags. The modification or addition of flags is limited to only one variable of the six at a time. First the variable to edit is selected, and the DQE highlights one or more data values for flagging by clicking and dragging. The selected values are highlighted and a list of possible flags appears (Fig. 12) to allow the DQE to select the appropriate flag(s). All flags are available for the DQE to add, including the preprocessor flags, but most flags added in this stage of the QC process are from the list in Table 7. In the following sections, the criterion for the use of each of the flags in Table 7 will be outlined.

Table 7: Additional QC flags assigned by DQE.

| Purpose | Flag |
|-------------------------------------|------|
| Discontinuity in data | H |
| Interesting feature in data | I |
| Data are erroneous - DO NOT USE | J |
| Data are suspect - USE WITH CAUTION | K |
| Known instrument malfunction | M |
| Spike in data | S |
| Data passed evaluation | Z |

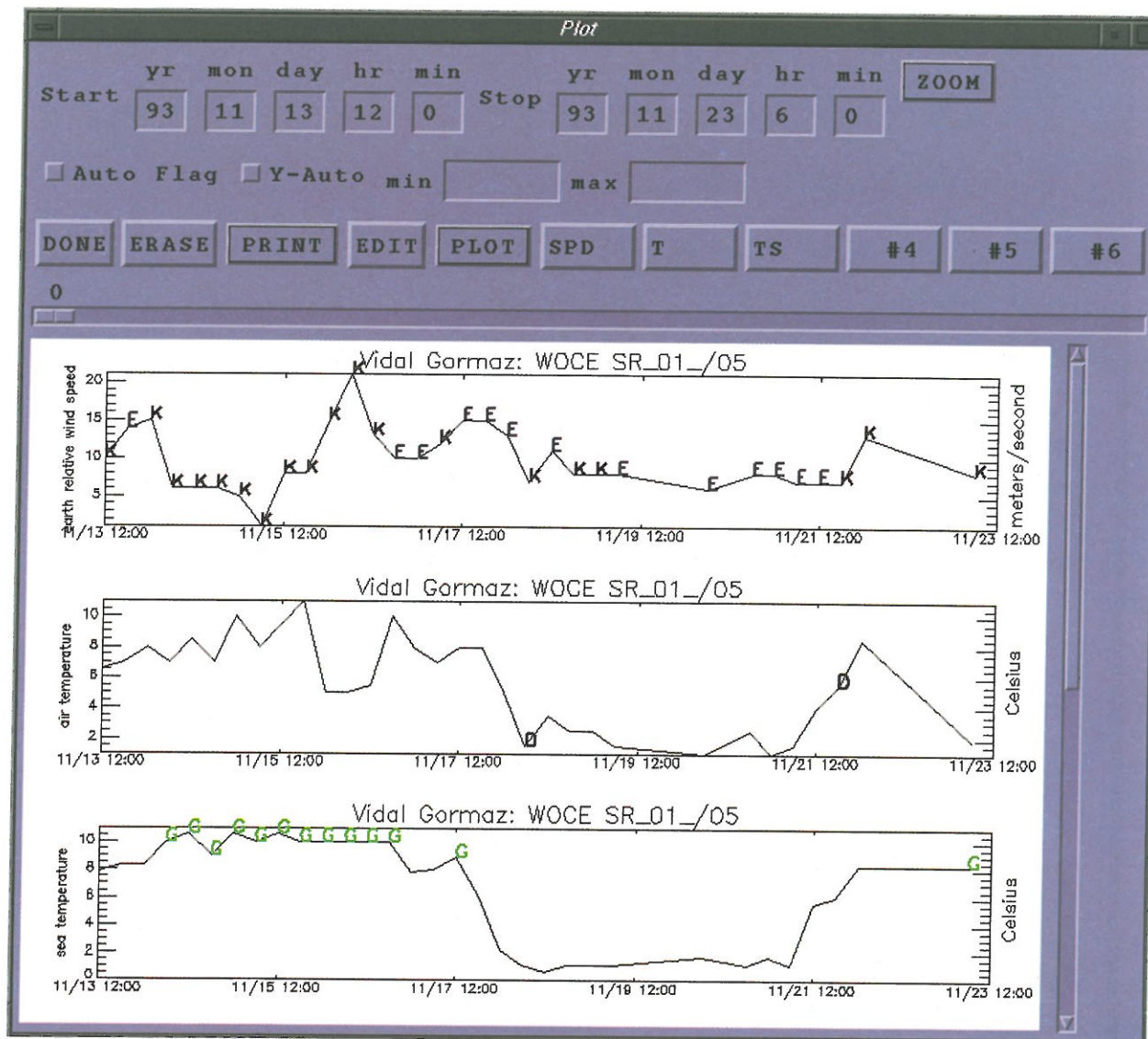


Figure 11. Screen shot of the VIDAT editor window with assigned QC flags displayed. Data for wind speed, air temperature, and sea temperature are shown for the 1993 SR-01 cruise of the R/V *Vidal Gormaz*

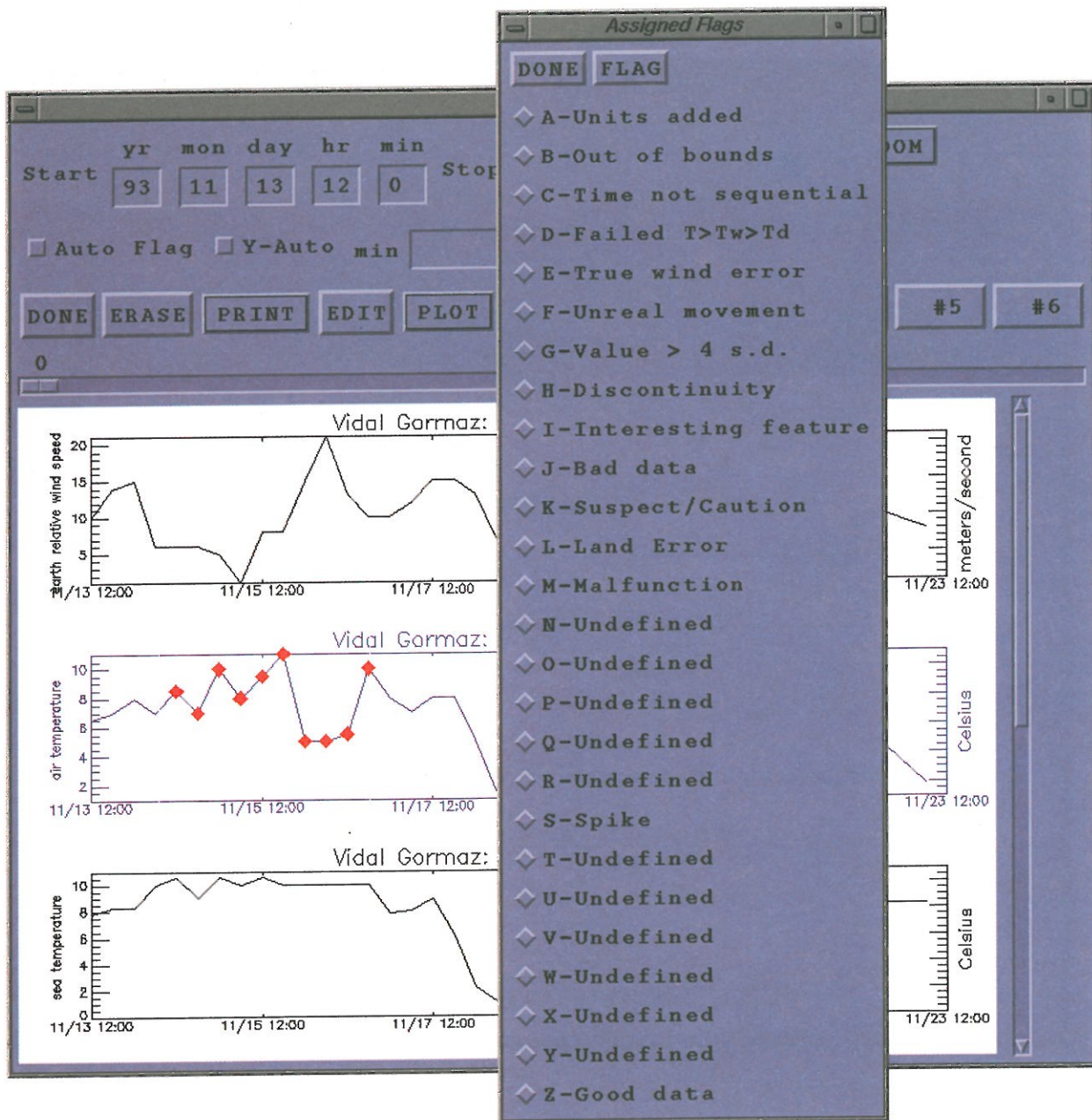


Figure 12. Screen shot of R/V *Vidal Gormaz* SR-01 data without flags displayed. The overlying window is a point and click flag selector used to flag the highlighted values.

Discontinuity

A discontinuity is defined as a sudden and dramatic shift in the data time series. The discontinuity often takes the form of a stair step. In the relative humidity data at Kapingamarangi on 17 November 1992 two discontinuities are noted (Fig. 13). For most of the day, and many of the preceding days, the relative humidity at Kapingamarangi held fairly steady between 70 and 80% (reasonable values for an island in the SW tropical Pacific Ocean). Around 1630 UTC on 17 November, the relative humidity dropped abruptly to near 30% and held near that value until 2100 UTC when the relative humidity abruptly returned to near 80%. The two sudden shifts in the relative humidity data are flagged using the "H" flag. Flags are placed at the beginning and ending points of the discontinuity (Fig. 13).

Discontinuities can occur for various reasons. A shift in sensor location or the replacement of one sensor with another can both cause discontinuities. However, a change of sensors or location may not result in a return of the values to their previous levels as exemplified in Fig. 13. Instead a stair step up or down in the data would occur. The cause of the discontinuity in Fig. 13 was a known sensor malfunction, which is discussed in the next section.

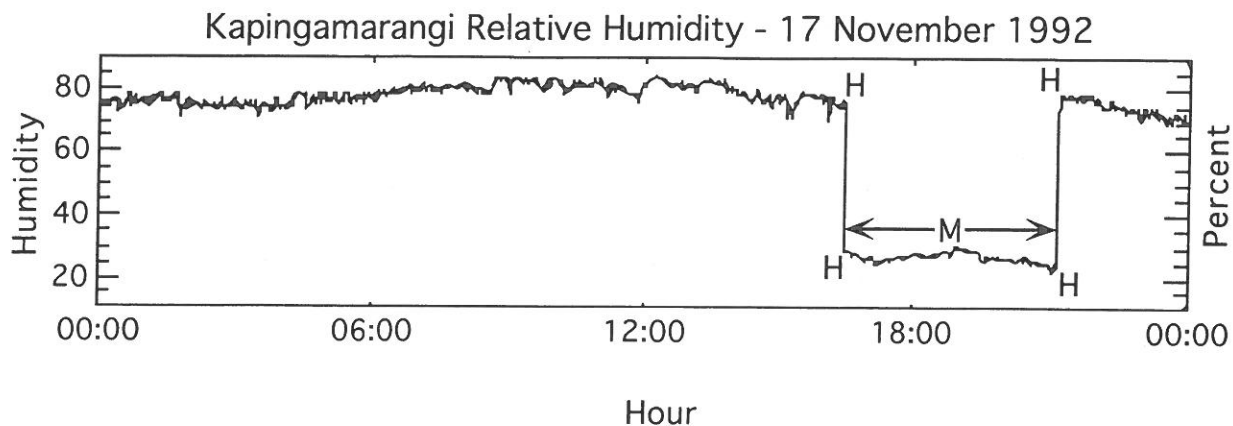


Figure 13. Discontinuity in the one-minute relative humidity data at the island of Kapingamarangi. The *H* flags are positioned at the starting and ending values of the discontinuity. In this case the discontinuity was caused by a sensor malfunction and all the values near 30% were flagged with an *M*.

Sensor malfunction

Sensor malfunctions are, as the name implies, a failure of the deployed sensor or data logging equipment. Malfunctions can occur for any number of reasons with breakage, poor calibration, improper maintenance, improper installation, power surges, and extreme weather events being just a few of the causes. The malfunction flag *M* is used only when supporting evidence of a sensor malfunction is available. Supporting information, i.e. metadata, is usually in the form of README files, or personal communication with the scientists that collected the meteorological data.

In the case of the relative humidity sensor at Kapingamarangi (Fig. 13), the data source informed the DAC that the sensor failed on 17 November 1992. The sensor functioned intermittently for the next few weeks; but on the 17th returned to normal operation at 2100 UTC. For the Kapingamarangi humidity data all values between 1630 and 2100 UTC were flagged as a malfunction.

Interesting feature

In the past, most flagging strategies only pointed out suspicious data within a set of meteorological data. Using a slightly different philosophy, the interesting feature flag was designed to mark unique features of meteorological data that have passed all other quality assurance tests. Values marked with the *I* flag should be considered valid, though they are often extreme. The interesting features and the variables that may be flagged in association with such an event, Table 8, are only flagged when either 1) there is evidence in more than one variable of an event or 2) metadata or independent data (e.g. satellite imagery) confirm the presence of such an event.

Table 8: Examples of variables that are often flagged for interesting features.

| Feature | Variable Flagged |
|---------------------|--------------------------|
| Hurricanes/Typhoons | Pressure |
| | Wind Speed |
| Convective Events | Temperature |
| | Relative Humidity |
| | Pressure |
| | Wind Speed |
| Frontal Passage | Temperature |
| | Wind Direction and Speed |

As an example, the atmospheric pressure and wind speed data for November 92 - February 93 at Nandi station in the western Pacific Ocean, Fig. 14, indicate that three tropical systems passed the station. Typhoons Joni, Kina, and Oli are marked as interesting features in all three cases for pressure and for the first two typhoons in the wind speeds. Less of a wind increase was seen for typhoon Oli; thus the wind speeds were not flagged. The interesting feature flag is placed at the extreme point that highlights the meteorological phenomenon. Whenever the interesting feature flag is used, a note describing the feature will appear in the QC report for the data set.

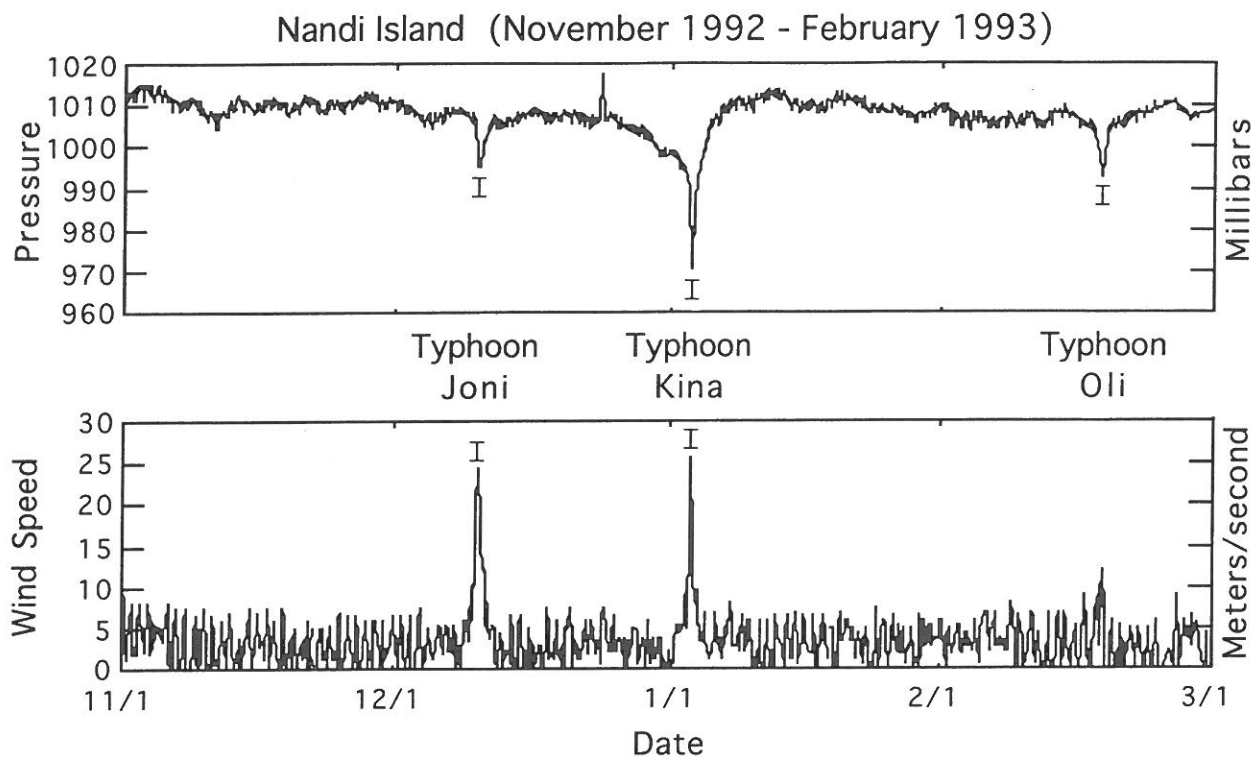


Figure 14. Typhoons Joni, Kina, and Oli passed the island of Nandi and were noted with the interesting feature flag *I*. Usually the *I* flag is applied to multiple variables, in this case atmospheric pressure and wind speed.

Erroneous data

While quality controlling data, values that are obviously in error are identified and assigned the erroneous data flag, *J*. The *J* flag marks data that are highly suspicious and SHOULD NOT BE USED. For example, Fig. 15 shows a four month period at Mili Atoll where the wind speed and direction were zero. That would imply that the wind did not blow for the entire four months! A truly unlikely occurrence at any location. Since we have no evidence to corroborate a sensor malfunction, these wind speed and direction data were flagged as erroneous.

In other cases, the *J* flag was used when metadata indicated the data were erroneous. When other information supports that the data are incorrect, the DQE flags the data so the end user can easily identify erroneous data.

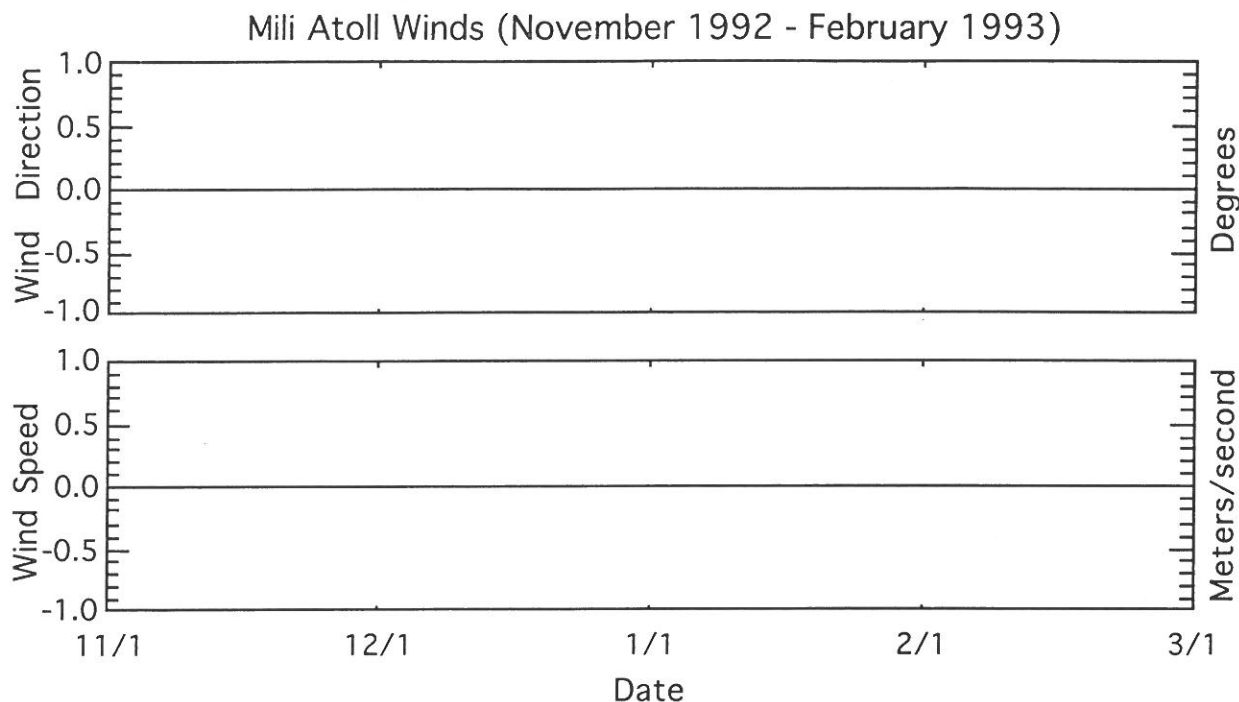


Figure 15. Wind speed and direction data from Mili Atoll were reported as zero from 01 November 1992 through 28 February 1993. All these highly unrealistic data were flagged as erroneous (*J*).

Suspect data

During the QC process, data values are often found that purely look suspicious or that do not fit the general trend of the remainder of the time series. Often there is no clear reason or supporting metadata to define these values as erroneous (J). These questionable data fall into a grey area, they are neither correct nor erroneous, just suspect, and are thus assigned a K flag to mark the data as suspect and urge that they be USED WITH CAUTION.

Some examples of suspect data are found on two different dates (14 February and 20 February 1993) on the island of Kapingamarangi, Fig. 16. Two different variables (atmospheric pressure and air temperature) show similar groups of suspect data values. The suspect regions in the data both follow gaps in the time series; the pressure (Fig. 16a) having lower than average values and the temperature (Fig. 16b) having higher than average values. No reason for the anomalies could be determined, and they were not flagged as discontinuities because they lack a stair step profile. As a result the data are flagged as suspect.

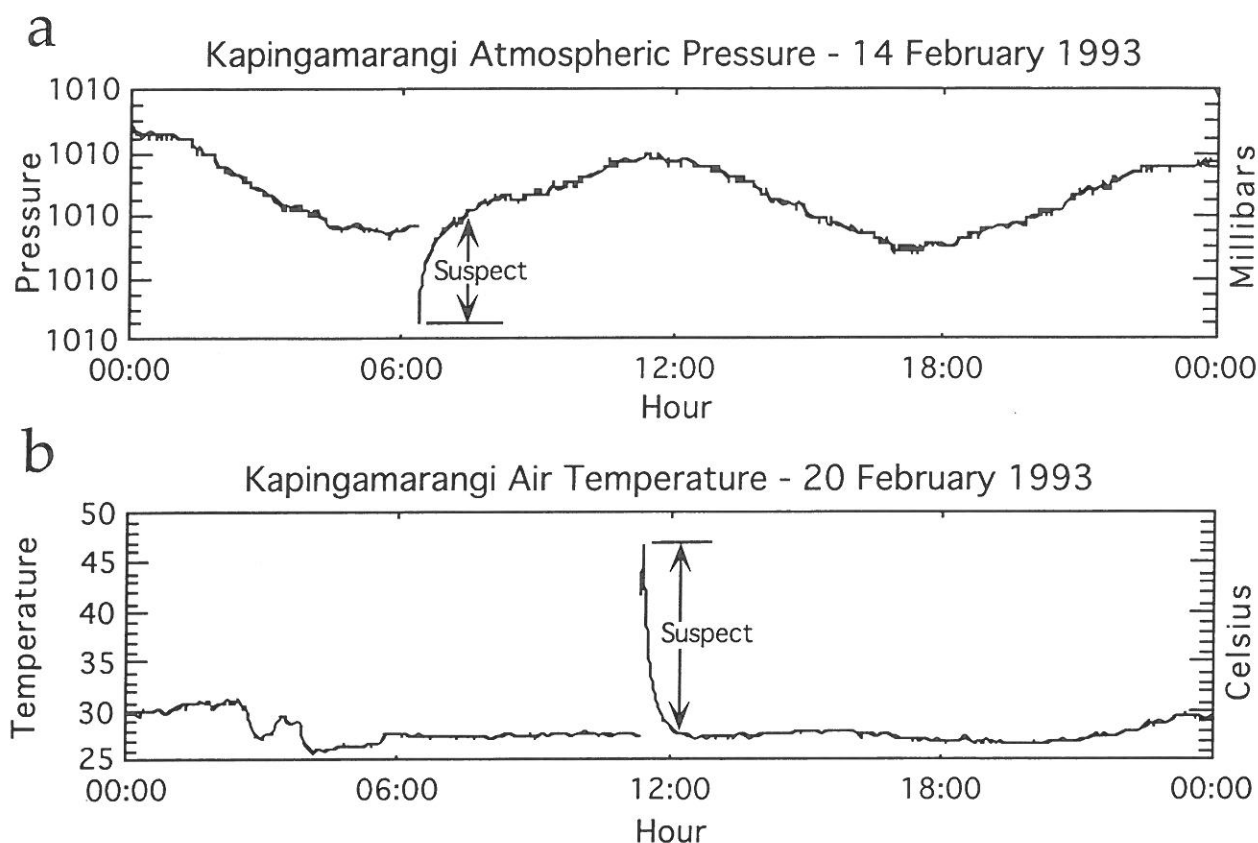


Figure 16. Examples of suspect pressure (a) and temperature data (b) on the island of Kapingamarangi in 1993. Regions between the arrows contain unlikely values that lacked supporting evidence for an instrument malfunction; thus they are flagged as suspect.

Spike in data

One commonality in any digitized meteorological data is the presence of spikes. Spikes are usually one or more data values that are singularly out of the general trend of the time series. For example a large spike in the wind speed data at Minamitorishima, Fig 17a, can be seen; the values increase from a normal of less than 15 m s⁻¹ to a peak of 50 m s⁻¹. With no changes in the other meteorological parameters to support such an increase, the value was flagged to indicate a spike (S).

Spikes are not necessarily large in amplitude. The Kapingamarangi pressure data contained a series of rhythmic, one millibar spikes, Fig. 17b. Though the actual cause of these spikes is unknown, speculation is they were caused by electronic interference from a neighboring instrument system. The DQE attempts to flag all spikes, regardless of magnitude.

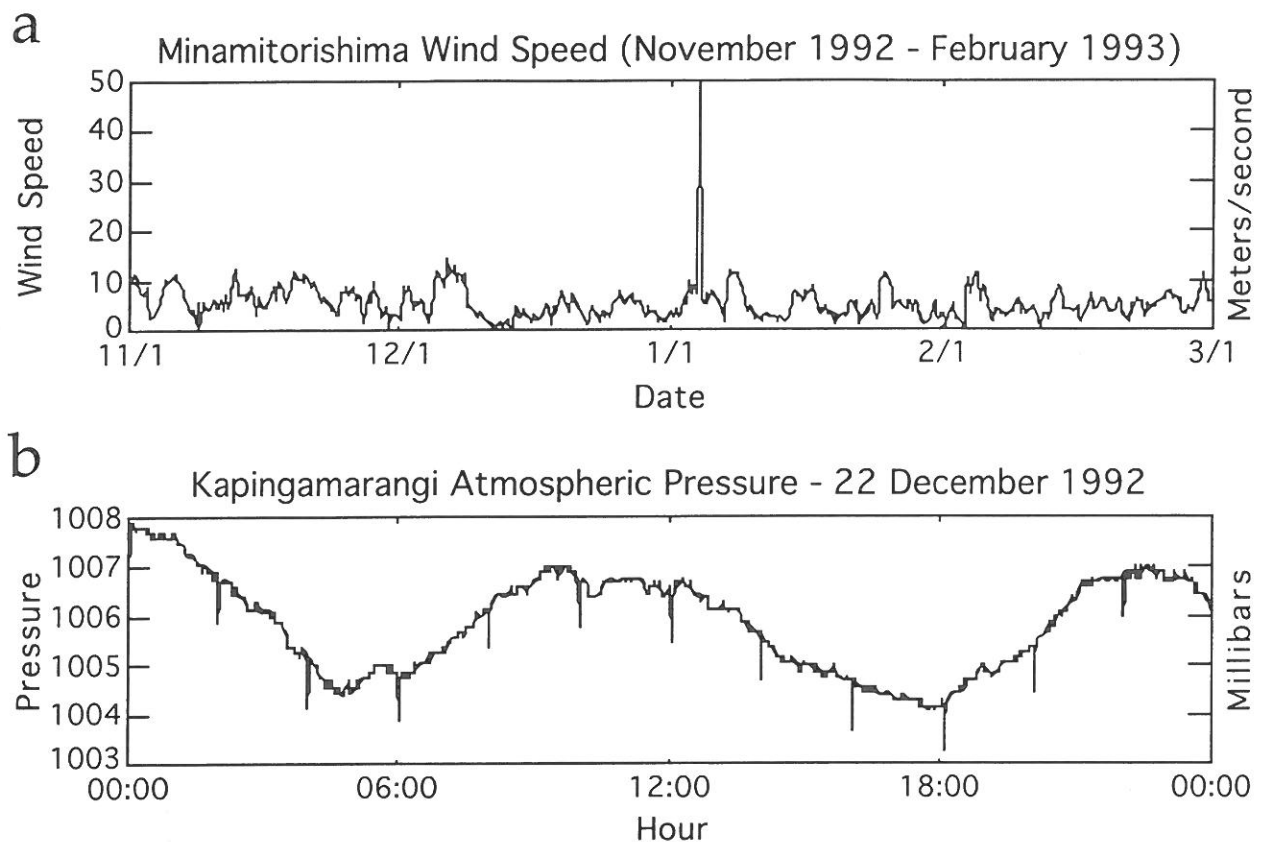


Figure 17. Examples of (a) a single large wind speed spike at Minamitorishima and (b) rhythmic one millibar atmospheric pressure spikes at Kapingamarangi. Both are flagged with an S.

Data passing evaluation

Finally data that passed all evaluation retain a Z flag. As was mentioned previously, all data values begin the evaluation process with the Z flag as we presume all incoming data are initially accurate until a value fails the preprocessor tests or visual inspection. The DQE can also reassign the Z flag to values flagged by the preprocessor if the DQE decides the data are valid.

4.4 New file creation and flag documentation

Once all desired flag changes have been made by the DQE, the DQE has the option of saving the file or reverting to previous versions. If the DQE selects to save the file, a new netCDF file is written combining the structure and data from the older netCDF file with the added flag changes. The FSU version number (Table 1) is automatically incremented and a new record is added to the history file. The history record contains the number of flag changes made for each variable, the date of the changes, and identifies the DQE.

When a netCDF file is saved and closed, a report file is produced. The report file contains another listing of the number of flags added to each variable. The DQE can add commentary to the report file. For example, the DQE places a comment line in the report that explains the "interesting feature" flags. Comments are also made whenever the DQE feels that his choice of flags will be unclear. The report files were designed to be used internally, however, DQE comments are reviewed and many are transferred to the QC reports.

7. REFERENCES

- da Silva, A. M., C. C. Young, and S. Levitus, 1994: Atlas of Surface Marine Data, Volume 1: Algorithms and Procedures. NOAA Atlas Series. In preparation.
- Smith, S.R. and D.M. Legler, 1995a: NetCDF Code Manual for Quality Controlled Surface Meteorological Data. Report WOCOMET 95-4, Center for Ocean Atmospheric Prediction Studies, Florida State University, Tallahassee, FL 32310
- Smith, S.R. and D.M. Legler, 1995b: ASCII Code Manual for Quality Controlled Surface Meteorological Data. Report WOCOMET 95-6, Center for Ocean Atmospheric Prediction Studies, Florida State University, Tallahassee, FL 32310
- Wallace, J.M. and P.V. Hobbs, 1977: Atmospheric Science, An Introductory Survey. Academic Press, Orlando, 467p.

Appendix A

Unidata netCDF, Version 2.4, February 1996

The Unidata network Common Data Form (netCDF) is an interface for scientific data access and a freely-distributed software library that provides an implementation of the interface. The netCDF library also defines a machine-independent format for representing scientific data. Together, the interface, library, and format support the creation, access, and sharing of scientific data. The current netCDF software provides common C and FORTRAN interfaces for applications and data. It has been tested on various common platforms, including several versions of UNIX, VMS, MSDOS, and OS/2.

NetCDF files are self-describing, network-transparent, directly accessible, and extendible. 'Self-describing' means that a netCDF file includes information about the data it contains. 'Network-transparent' means that a netCDF file is represented in a form that can be accessed by computers with different ways of storing integers, characters, and floating-point numbers. 'Direct-access' means that a small subset of a large data set may be accessed efficiently, without first reading through all the preceding data. 'Extendible' means that data can be appended to a netCDF data set without copying it or redefining its structure.

NetCDF is useful for supporting access to diverse kinds of scientific data in heterogeneous networking environments and for writing application software that does not depend on application-specific formats. A variety of analysis and display packages have been developed to analyze and display data in netCDF form.

You can obtain a copy of the latest released version of netCDF software using a WWW browser or anonymous FTP from

`ftp://ftp.unidata.ucar.edu/pub/netcdf/netcdf.tar.Z`

Included in this distribution are: the C source for the netCDF data access library, sources for the FORTRAN jacket library for various systems, documentation for the netCDF library and utilities in the form of a netCDF User's Guide, source for the netCDF utilities `ncdump` and `ncgen`, a directory of test programs to verify the correct implementation of the netCDF library in new environments, and a directory of XDR (eXternal Data Representation) source code for environments that do not support XDR.

Other files about netCDF are available from the URL

`http://www.unidata.ucar.edu/packages/netcdf/`

and include:

| | |
|------------------|--|
| README | • general information about netCDF. |
| FAQ | • Frequently Asked Questions (with answers) about netCDF. |
| utilities.txt | • a list of software packages currently available or under development for manipulating and displaying netCDF data. |
| guide.ps.Z | • a compressed PostScript file of the NetCDF User's Guide. This is included in the netcdf.tar.Z distribution, so you don't need both. |
| ncprogs.ps | • a draft PostScript document describing an initial set of netCDF operator and utility programs under development. |
| ncprogs.txt | • an ASCII version of ncprogs.ps. |
| conventions.info | • a draft document of some proposed netCDF conventions. |
| cdl/ | • a directory containing some examples of CDL files (an ASCII representation for netCDF files). |
| msdos/ | • a directory containing executables and binaries for netCDF under MSDOS 5.0. These can also be built from the sources in netcdf.tar.Z, if you have the necessary Microsoft compilers. |
| mac/ | • a directory containing notes and Macintosh MPW makefiles for porting netCDF to an Apple Macintosh. These were contributed by Chuck Denham, U.S. Geological Survey. |

A mailing list, netcdfgroup@unidata.ucar.edu, exists for discussion of the netCDF interface and announcements about netCDF bugs, fixes, and enhancements. For information about how to subscribe, see the URL

<http://www.unidata.ucar.edu/packages/netcdf/mailling-lists.html>

An archive of past postings to the netcdfgroup mailing list is available for searching from the netCDF home page.

A recent paper that provides a good introduction to the use of netCDF appeared in

Jenter, H. L. and R. P. Signell, 1992. "NetCDF: A Freely-Available Software-Solution to Data-Access Problems for Numerical Modelers". Proceedings of the American Society of Civil Engineers Conference on Estuarine and Coastal Modeling. Tampa, Florida.

This paper is available via anonymous FTP from

host: crusty.er.usgs.gov
file: pub/netcdf.asce.ps

Specific questions about netCDF that are not of interest to the netcdfgroup mailing list may be sent to support@unidata.ucar.edu.

Appendix B

Sample listing of the contents of a public netCDF file created by the FSU DAC. This file is stored in a binary format but the listing presented here can be created using a netCDF utility called "ncdump" (refer to information provided by Unidata, Appendix A).

```
netcdf CCVG.931007011v100 {
dimensions:
    rec = 43 ;
    f_string = 12 ;
    ctc_string = 9 ;

variables:
    char ctc(rec, ctc_string) ;
        ctc:long_name = "cruise track code" ;
        ctc:FORTTRAN_format = "A9" ;
    long time(rec) ;
        time:units = "minutes from 1-1-1980 00:00" ;
        time:type = 2 ;
        time:ave_period=0;
        time:ave_center=0;
        time:qcindex = 1 ;
        time:FORTTRAN_format = "I12" ;
    float lat(rec) ;
        lat:long_name = "latitude" ;
        lat:units = "degrees" ;
        lat:convers_units = 0 ;
        lat:qcindex = 2 ;
        lat:FORTTRAN_format = "F9.2" ;
    float lon(rec) ;
        lon:long_name = "longitude" ;
        lon:units = "degrees east" ;
        lon:convers_units = 0 ;
        lon:qcindex = 3 ;
        lon:FORTTRAN_format = "F9.2" ;
    float PL_HD(rec) ;
        PL_HD:long_name = "platform heading" ;
        PL_HD:units = "degrees (clockwise from true north)" ;
        PL_HD:convers_units = 5 ;
        PL_HD:inst = "Magellan 5000D GPS" ;
        PL_HD:qcindex = 4 ;
        PL_HD:FORTTRAN_format = "F9.0" ;
    float PL_SPD(rec) ;
        PL_SPD:long_name = "platform speed" ;
        PL_SPD:units = "meters/second" ;
        PL_SPD:convers_units = 5 ;
        PL_SPD:inst = "Magellan 5000D GPS" ;
        PL_SPD:qcindex = 5 ;
        PL_SPD:FORTTRAN_format = "F9.1" ;
    float DIR(rec) ;
        DIR:long_name = "earth relative wind direction (meteorological)" ;
        DIR:units = "degrees true" ;
        DIR:convers_units = 0 ;
        DIR:ht = 15.24 ;
        DIR:inst = "calculated from F420G Electric Speed Indicator (USA)" ;
        DIR:qcindex = 6 ;
        DIR:FORTTRAN_format = "F9.0" ;
    float SPD(rec) ;
        SPD:long_name = "earth relative wind speed" ;
        SPD:units = "meters/second" ;
        SPD:convers_units = 5 ;
        SPD:ht = 15.24 ;
        SPD:inst = "calculated from F420G Electric Speed Indicator (USA)" ;
        SPD:qcindex = 7 ;
        SPD:FORTTRAN_format = "F9.0" ;
```


Appendix C

Definitions of WOCE quality control flags and reference page numbers in this report detailing their use.

| Flag | Definition | Page |
|------|--|------|
| A | Original data had unknown units. The units shown were determined using a climatology or some other method. | 10 |
| B | Original data were out of the range bounds outlined (Table 5). | 14 |
| C | Time data are not sequential or date/time not valid. | 12 |
| D | Data failed $T \geq T_w \geq T_d$ test. In the free atmosphere, the value of the temperature is always greater than or equal to the wet-bulb temperature, which in turn is always greater than or equal to the dew point temperature. | 23 |
| E | Data failed resultant wind recomputation check. When the data set includes all variables required, a program recomputes the earth relative wind speed and direction and compares the computed values to the reported earth relative wind speed and direction. A failed test occurs when the wind direction difference is $> 10^\circ$ or the wind speed difference is > 5 m/s. | 20 |
| F | Platform velocity unrealistic. Determined with platform position data. | 16 |
| G | Data are > 4 standard deviations from the COADS climatological means (da Silva et al. 1994). Test applied only to pressure, temperature, sea temperature, relative humidity, and wind speed. | 12 |
| H | Discontinuity found in data | 33 |
| I | Interesting feature found in data. More specific information on the feature is contained in the QC reports. Examples include: hurricanes, sharp sea water temperature gradients, strong convective events, etc. | 34 |
| J | Data are of poor quality by visual inspection, DO NOT USE. | 36 |
| K | Data suspect - USE WITH CAUTION - flag applied when the data look to have obvious errors, but no specific reason for the error can be determined. | 37 |
| L | Oceanographic platform passes over land. | 17 |
| M | Known instrument malfunction. | 34 |
| P | Position of platform or its movement are uncertain. Data should be used with caution. | 10 |
| Q | Data arrived at the DAC already flagged as questionable. | 10 |
| S | Spike in the data. Usually one or two sequential data values (sometimes up to 4 values) that are drastically out of the current data trend. Spikes occur for many reasons including power surges, typographical errors, data logging problems, lightning strikes, etc. | 38 |
| T | Time duplicate | 12 |
| Z | Data passed evaluation | 39 |