DBCP Buoy Monitoring Statistics

Table of Contents

Introduction

Buoy monitoring statistics are produced (or have been produced) monthly by following monitoring centres:

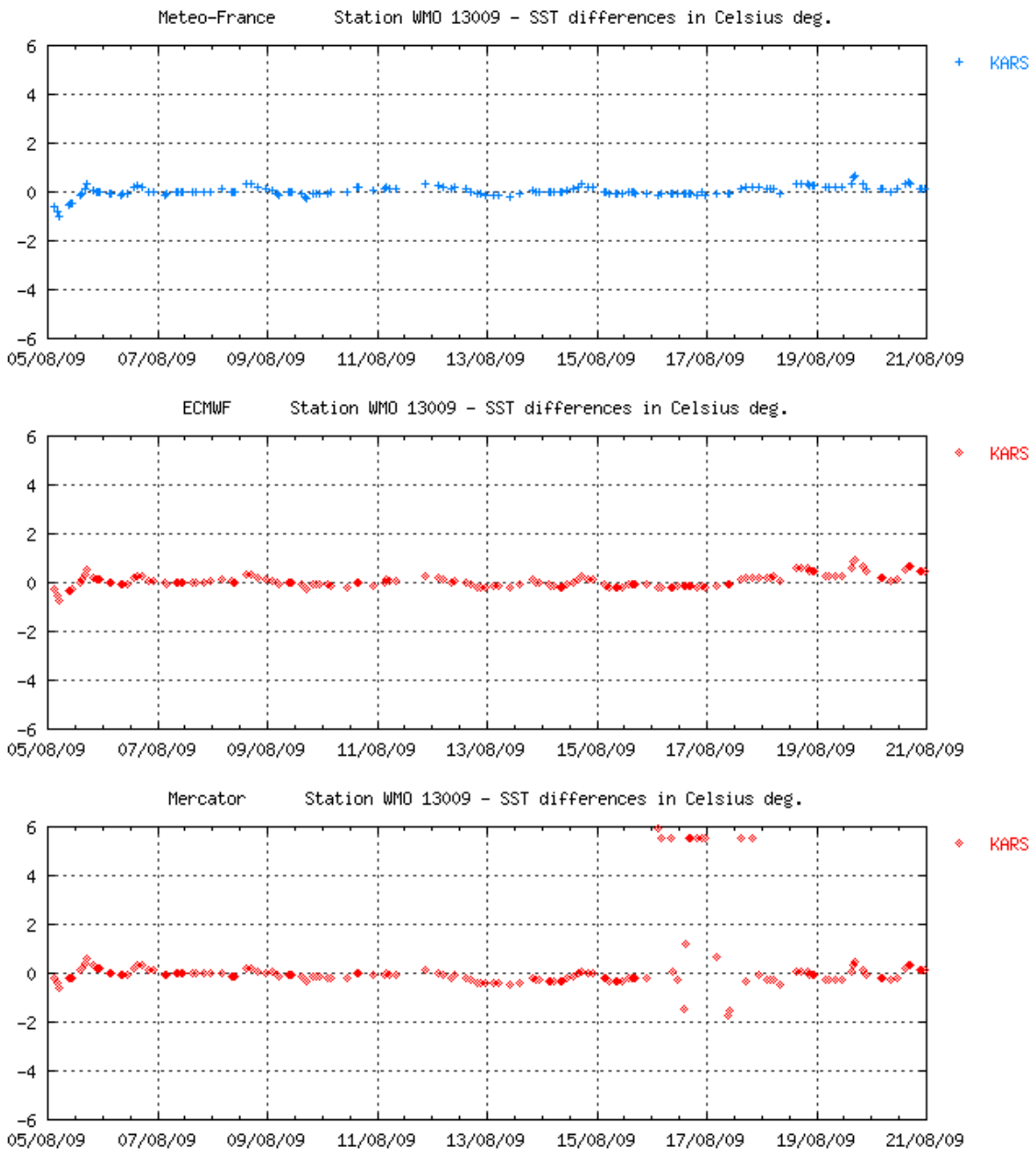
1. CMM: Centre de Météorologie Marine of Météo France, Brest, France;
2. NCO: NOAA National Center for Environmental Prediction (NCEP) Central Operations, Suitland, Maryland, USA;
3. UKMO: United Kingdom Meteorological Office, Bracknell, UK;
4. ECMWF: European Center for Medium Range Weather Forecasts, Reading, UK.
5. SAWB: South African Weather Bureau, Pretoria.
6. BOM: Australian Bureau Of Meteorology

Buoy data for a number of geo-physical variables (e.g. Air Pressure) collected from the Global Telecommunication System (GTS) are compared with the first guess fields (i.e. 6 hour forecast) or analysis fields of the main global model of each centre. Normally all statistics are computed excluding the gross errors. Gross errors are defined as observations departing from

the background field (i.e. first guess field) by more than a constant value (e.g. 15 hPa for air pressure).

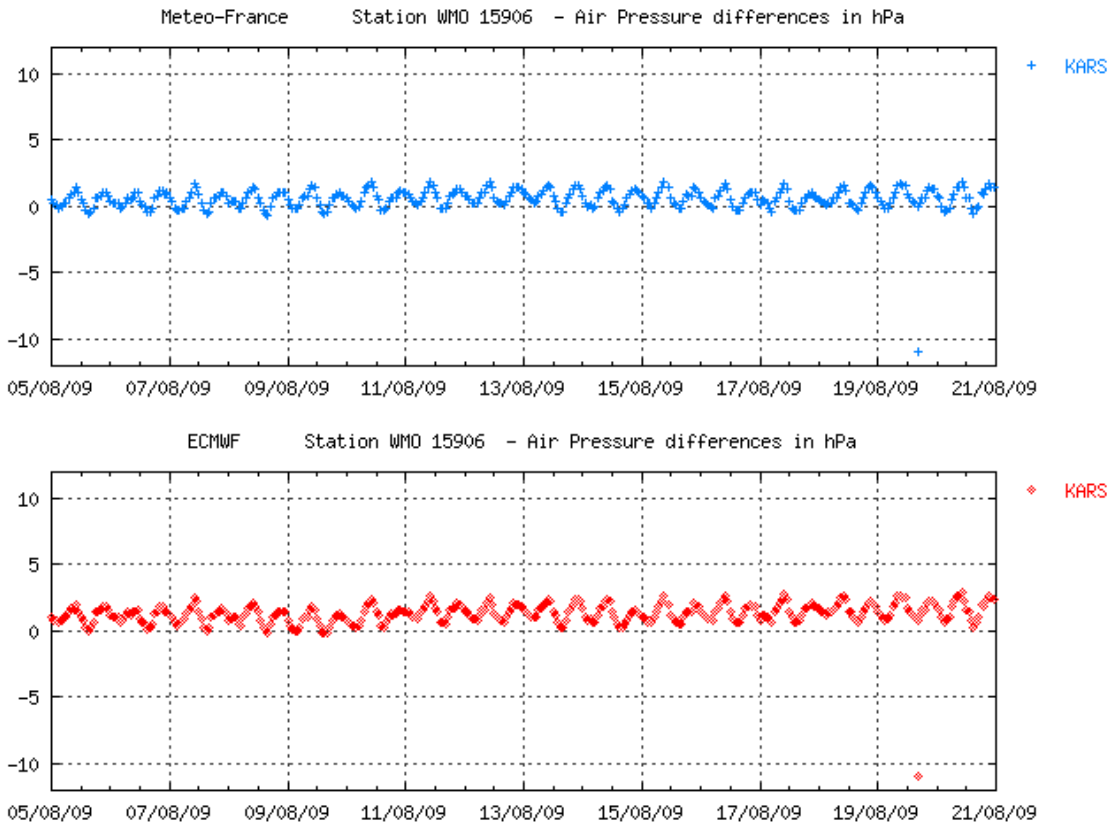
Comparison of observed data with model first guess field is being considered at the monitoring centres as an excellent method for estimating the quality of the observations.

The figure below shows difference observation minus first guess of Météo France model for air pressure for buoy 44762 in April 1998. Bias of about -3 hPa appears on the graph:



Example of QC Plot for a mooring (Sea Surface Temperature compared to various models)

http://www.meteo.shom.fr/cgi-bin/meteo/dataplot_surfmar.cgi?wmo=13009&type=0&sns=sst



Example of a QC plot for a drifter. Air pressure compared to two models.
http://www.meteo.shom.fr/cgi-bin/meteo/dataplot_surfmar.cgi?wmo=15906+&type=0&sns=ap

Buoy monitoring statistics produced by Météo France for this buoy give following figures for April 1998 which confirm (standard deviation is small) a bias between -2.5 and -3 hPa:

GTS source	# Obs.	Bias	SD	RMS
BGSF	23	-2.5	0.5	2.5
ENMI	160	-2.9	0.9	3.0
LFPW	43	-2.6	0.7	2.7

For example, distribution of (Observation - First guess) is nearly Gaussian for the vast majority of the marine observations.As shown on the example above, the few buoys reporting bad (large SD) or biased (large bias, small SD) data will easily be detected that way.There are ocean area like in the southern hemisphere where the first guess field is not as good as

expected and where higher values of RMS might be due to the model itself, not to the observations. This is due in part to the lack of observed data in those area. Monitoring centres take this into consideration and use higher limits for RMS in those area for alerting user community on problems with buoys.Problems are reported through the DBCP QC Guidelines.

Since analysis is not independent from the observations, comparison of observed data with the first guess field is preferably chosen. There are however cases where first guess field is not available (e.g. SST) while analysis is available. In that case, comparison with the analysis can be useful to estimate the quality of the data although other tools would have to be used to go further (e.g. comparison with other nearby observations).

Not all observations received at each centre are used and inserted into the model. Since models and data assimilation schemes are different, buoy monitoring statistics are not consistent among the different centres. Time windows, cut off times, and real time quality control procedures (e.g. black listing) leads to certain data being rejected by a model and are not therefore included in the statistics. For example, number of observations accepted by the model may vary widely from one model to the other. Standardization effort has been undertaken to harmonize the statistics but certain differences will always remain. Hence although standard proposed below is followed by all centres as far as format is concerned it is not necessarily followed as far as algorithms, gross errors, etc. are concerned.

Standard format for the buoy monitoring statistics:

Distribution of the statistics:

Statistics are produced monthly by the four centres and distributed onto an Internet mailing list. To register on the mailing list, and receive the buoy monitoring statistics automatically, send a message containing "SUBSCRIBE dbcp-qc" (or "UNSUBSCRIBE dbcp-qc") in the subject of the message to the address sympa@sympa.cls.fr.

Statistics are also archived by Météo France and made available via the web:

- ftp://esurfmar.meteo.fr/pub/pb/qc-stats/ (replaces ftp://ftp.shom.fr/pub/meteo/qc-stats) : To download monthly files
- http://wo.jcommops.org/cgi-bin/WebObjects/JCOMMOPS.woa/wa/statsSeries?prog=DBCP: Query form to access statistics for one or more buoys
- http://wo.jcommops.org/cgi-bin/WebObjects/JCOMMOPS.woa/wa/histogram?prog=DBCP: Produce a chart of Monitoring Stats
- http://www.shom.fr/meteo/rechstat: Simple query form to access statistics for an individual buoy

Through the mailing list, subject line for buoy monitoring statistics is standardized as follow:

STAT center ppp year mm dd

STAT is constant, i.e. the four letters "STAT"

color="#008000">**center**: Name of the center producing the statistics, e.g.

- ECMWF: European Center for Medium Range Weather Forecasts
- NCO: NOAA NCEP Central Operations
- CMM: Meteo France, Centre de Meteorologie Marine
- UKMO: United Kingdom Meteorological Office
- SAWB: South African Weather Bureau
- SMN: Servicio Meteorológico Nacional, Argentina
- BOM: Bureau Of Meteorology, Australia

ppp: Type of physical variable concerned or ALL if many variables are included:

- AP: Air Pressure (coded as 'AP ')
- AT: Air Temperature (coded as 'AT ')
- SST: Sea Surface Temperature
- WD: Wind Direction (codes as 'WD ')
- WS: Wind Speed (coded as 'WS ')
- WV: Wind vector (i.e. u and v components)
- APT: Air Pressure Tendency
- POS: Position of the buoy
- TZ: Subsurface temperatures (coded as 'TZ '): The depths of the probes and proposed actions should be placed in the body text, not in the subject line (not enough room)
- ALL: All buoy sensors (e.g. remove all buoy data from GTS)

year:Year concerned (e.g. 1998)

mm: Month concerned (e.g. 04 for April)

dd: Last day of the 1-month period concerned. It is optional and used only if the 1-month period does not end on the last day of the month. For example dd=15 if the 1-month period concerned is 16 July to 15 August.

Example:

```
From                    Date            Subject

dbcp-qc@jcommops.org      02-May-1998    STAT CMM ALL
1998 04
```

The April 1998 monitoring statistics for many geo-physical variable and produced by the Centre de Meteorologie Marine of Meteo France are available in the body text.

Standard format for the buoy monitoring statistics:

Format of the statistics:

The monitoring statistics are available in the body text. Header part contains information upon format and gross errors used. There is one line of data per buoy per geo-physical variable. Format is fixed and fields are separated by commas ','. Missing values appear as blanks. Fields appear in the following order:

Date####: Oldest date between: last day for the monthly statistics and last day of data reception on GTS for the monitored sensor (format YYYYMMDD)

WMO##: WMO number of buoy or Ship's Call Sign

Sns: Sensor Name :

- AP: Pressure
- AT: Air Temp
- SST: Sea Surf Temp
- WS: Wind Sp
- WD: Wind Dir
- WV: Wind vector
- APT: Tendency
- HUM: Humidity
- TD: Dew Point

Orig: GTS Origin of the data (ALL or cccc from GTS Bulletin header)

C: GTS code (B: BUOY, S:SHIP, Y:SYNOP)

Cntr#: Monitoring Center producing the stats (e.g. ECMWF, UKMO, OPC, CMM)

Lat##: Last Latitude of buoy/ship during the month

Long##: Last Longitude of buoy/ship during the month

Rcei: Total number of obs received at the center including obs not used

Acpt: Total number of obs accepted by the model

GE#: Number of Gross Errors (i.e. number of (Obs-Field) exceeding limits)

Bias#: Mean Bias (Obs-Field)

Special algorithm for bias of wind direction only: For buoys reporting wind direction with an offset close to 180 degrees, usual algorithm would show a very large SD and bias making it difficult to detect the error. Ideally we should have a small SD with a bias in the order of 180 degrees. This is possible by applying the following algorithm presently used by CMM:

- Compute the 4 series modulo 360 (i.e. result should fit in range -180 .. +180) of SD and bias for

1) (Obs-FG)

2) ([Obs-90]-FG)

3) ([Obs+90]-FG), and

4) ([Obs+180]-FG)

- In each series, remove gross errors, e.g. values of |[Obs-180]-FG| >= GE limit: if there is a 180 degrees bias then [Obs-180]-FG will probably fit within the limits.
- Then pick the smallest SD of the 4 series as the actual SD, and take (bias-i) as the actual bias if series ([obs+i]-FG) was picked as the one with the smallest SD.

SD##: Standard Deviation, SD = RMS (Obs-Field-Bias);

Computation of SD for wind vectors only: By convention a special algorithm is applied: SD=RMS(WS/Rate - Field).

Remark: As an exception, it is understood that values of SD and Rate given for Wind Vectors are actually computed based upon Wind Speed, not Wind Vector.

RMS#: Root Mean Square, RMS = RMS (Obs-Field);

For Wind Vectors (WV), by convention, RMS=RMS(SQRT(Vec(WV-Field)**2))

Rate: Mean (Obs/Field)

Computation of Rate for wind vector and only for wind vector: Mean(Obs./Field).

Remark: As an exception, it is understood that values of SD and Rate given for Wind Vectors are actually computed based upon Wind Speed, not Wind Vector.

F: Flag for Field used : A=Analysis, G=First Guess, B=Both

Example:

Date####,WMO##,Sns,Orig,C,Cntr#,Lat##,Long##,Rcei,Acpt,GE#,Bias#,SD##,RMS#,Rate,F

19980430,12511, AP,KARS,B, CMM, 27.7, 35.5, 657,   , 0, -2.6, 1.0, 2.8,   ,B

19980430,12511, AT,KARS,B, CMM, 27.7, 35.5, 620,   , 0, 2.4, 3.2, 4.0,   ,B

19980414,12511,SST,KARS,B, CMM, 27.7, 35.5, 228,   , 56, 1.5, 2.1, 2.6,   ,B

19980409,12512, AP,KARS,B, CMM, 24.1, 35.5, 271,   , 0, 0.8, 0.8, 1.2,   ,B

19980409,12512, AT,KARS,B, CMM, 24.1, 35.5, 271,   , 0, -1.7, 2.7, 3.2,   ,B<

Specific settings and tests for the statistics produced by the different centres

	UKMO	CMM	ECMWF	NCO	SAWB	BOM	Suggested
AP	15hPa	10hPa	15hPa	15hPa	15hPa	15 hPA	15hPa
AT	15C	15C	10C	15C	/	/	15C
WS	15m/s vect.	15m/s	15m/s	15 m/s	/	15 m/s	15 m/s
WD	100 Deg	100 Deg	100 Deg	100 Deg	/	100 Deg	100Deg
WV	25m/s	?	25 m/s	25m/s	/	25 m/s	25m/s
SST	/	5C	/	5C	/	5C	5C

Table 1: Gross error limits used by the different centres:

ECMWF

- Rcei column indicates the number of observations that are presented to the model. Observations arriving after the cut off time (unknown value) are not presented to the model.
- Statistics are computed based on the 4D variational data assimilation scheme.
- Acpt column indicates number of observations used by the model (not counting the gross errors nor observations out of time window (value unknown)).
- Only one observation is used for each cycle. Other observations are ignored and are not even used for computing the statistics. 4 cycles are run per day at 00UTC, 06UTC, 12UTC, and 18UTC.
- The statistics are computed using the feedback files generated by the data assimilation scheme. Air temperature and SST coming from Drifters are not presented to the model so the computation of such statistics is not possible.
- **Things that might be proposed in the future for more consistency:**

1. Compute specific SD algorithm for Wind Vector (WV)
2. Compute special algorithm for bias of Wind Direction (WD)
3. Indicate what is the cut off time and time windows which are used. Indicate whether those are applied to the Rcei and/or Acpt columns.

UKMO

- The UKMO data assimilation system is a 3DVar scheme. It treats all observations as if valid at the analysis time. In the Incremental Analysis Update initialization scheme, the analysis increments are added gradually to the model during integration from 3 hours before until 3 hours after the analysis time.
- For the global model which is used for computing the buoy monitoring statistics, observations within 3 hours of the nominal analysis time are used with a cut-off of 3 hours 20 minutes.

  For example, for the 1200 UTC global analysis, model uses observations that are valid between 0901 and 1500 TC, and that have arrived before 1520 UTC. The global model 6-day forecast then runs. If late observations are received after 1520 UTC but before 2320 UTC then they will be used in the 'update' analysis, which is used to provide a background field for the following midnight analysis.

  Data assimilation scheme 'thins' frequent observations from the same buoy so that it only uses a maximum of 5 reports for a global analysis, or 3 for a limited area analysis. In the case of the global model, the reports are sorted by observation time minus analysis time (in minutes) into intervals of:

    - -180 to -105, -105 to -35, -35 to 35, 35 to 105 and 105 to 180.

  Within each bin, those observations closest to -140, -70, 0, 70 and 140 respectively are selected unless they are within 60 minutes of the previous report.

NCEP

The following description of GDAS assimilation scheme is as of August 1998. The Global Data Assimilation System (GDAS) runs four times a day. Table below shows the approximate start times for each cycle. Each GDAS run will use data with observation times +/- 3 hours from the "on" time of the GDAS run. Following table also shows the data time windows for each GDAS run:

Run	Run begins at	Time window
0000 GDAS	0600	2100 (previous day) to 0300
0600 GDAS	0940	0300 to 0900
1200 GDAS	1950	0900 to 1500
1800 GDAS	2200	1500 to 2100

All data that are received up until the start time of each GDAS run may be used, as long as that data have observation times within the +/- 3 hour time window for that GDAS run. For example, if a report with an observation time of 1100Z is received at 1930Z, it may be used in the 1200Z run of the GDAS (which begins around 1950Z), depending on the quality of the report. In addition, all data must include a pressure observation in order to be used. If any surface marine report is missing both sea-level pressure and surface pressure, the report will not be used by the GDAS. If either sea-level pressure or surface pressure are included in an observation, that observation may be used by the GDAS. The only exception to this rule

concerns observations from ATLAS buoys. An ATLAS buoy report that is missing a pressure observation will have the standard atmosphere value of 1013.25 mb attached to the report, so that other parameters reported by the ATLAS buoy may be used by the GDAS.

The GDAS will use all reports received that have good or neutral quality marks. These quality marks may have been set by manual quality control or by an automated process. For surface marine data, the quality marks checked by the GDAS are those set on surface pressure (same as sea-level pressure for marine platforms at sea-level), air temperature, and wind. No thinning is done for multiple observations from any one platform. All observations have the same weight, no matter what their observation time is (i.e. for the 1200Z GDAS run, a 1200Z observation has no more weight than a 1000Z observation or a 1400Z observation). In order to use off-time observations, the GDAS takes forecasts valid at the "on" time and at +/- 3 hours and performs a time interpolation between the forecasts and the observation. The differences computed between the interpolated forecasts and the observation are then applied to the GDAS. For example, if a buoy report has an observation time of 1000Z, then forecasts valid at 0900Z and 1200Z would be used and interpolated in time to 1000Z. Differences would be computed between the interpolated forecast and the 1000Z buoy report, and those differences would then be applied to the 1200Z GDAS.

- **Things that might be proposed in the future for more consistency:**
    - Adding information in column Acpt for SST

CMM

- **Things that might be proposed in the future for more consistency:**
    - CMM: CMM is not in a position of include number of accepted observations because the information is not available. CMM could however simulate data assimilation scheme of the model in order to come with close enough figures. CMM is presently studying feasibility.

SAWB.

- No longer active
- Statistics excludes Gross errors; Limit for AP: 15.0 mb

  Numbers of observations refer to the statistics - NOT the NWP model.

  All statistical measures are weighted according to proximity of observation time to model reference hour; Time range is +/- 2 hours.

  Buoys in the window 50W to 70E ; Equator to 60S.

- SAWB uses ECMWF Analysis. The model fields used are the GRIB files HPIA98 and HPLA98 as on the GTS with some identification parameters the following: Center = 98, model/process = 190. The ECMWF Analysis is only valid for 12 UTC.

  Are accepted all buoy observations (Obs) in the time interval (10,00 - 14,00) UTC for evaluations using ECMWF.

  The actual position of each Obs is interpolated into the model field to obtain the model value at that point. The difference between Obs and Model value is found and saved in an unique file for each buoy. These differences are accumulated for a whole month.

  Obs differences are weighted according to the distance from 12 UTC in the following manner:

  an Obs at 12 gets a weight of 9 (=3X3)
  an Obs at 11 or 13 is weighted with 4 (=2x2)
  an Obs at 10 and 14 has a weight of 1 (=1x1)
  and an Obs at say 10:30 (=10.50) gets a weight [(3-1.5)**2 = 2.25] rounded to 2.

  Gross errors are excluded from calculating statistics.

  After the end of a month all the accepted Obs (differences) with their weights are taken together for calculating the statistics. The formulae used are:

  BIAS = [SUM(Wi x Di)] / [SUM(Wi)] ; Wi = weight(i) and Di = Difference(i)

  RMS = SQRT{ SUM[Wi x (Di **2)] / SUM(Wi) }

  SD = SQRT{ 1/[SUM(Wi)-1] x SUM[Wi x (Di **2)] -

  [SUM(Wi)] / [SUM(Wi)-1] x [BIAS]**2 }