Quality control of ocean temperature and salinity profiles –

historical and real-time data

Bruce Ingleby and Matt Huddleston,
Met Office, FitzRoy Road, Exeter, EX1 3PB, UK

Final version - November 2005
To appear in Journal of Marine Systems

bruce.ingleby@metoffice.gov.uk

Abstract

A new automated quality control system for oceanic temperature and temperature-salinity profiles is presented. Substantial development was needed for some of the quality control algorithms although the checks were based on documented procedures used elsewhere where appropriate. A new automated ship track check was developed: the results of an undetected position error can be very damaging to ocean analyses. Also important is a check against a gridded background, this can be a climatology but near the surface it is advantageous to use an estimate that is evolving over time. Bayesian probability theory is used in the background check and the associated check against nearby observations (buddy check). The system was used to process archive data for 1956-2004. As a by-product monthly model-free objective analyses for this period were produced. Versions of the system are used for near-real time ocean analysis and for initialising both short-range ocean forecasts and seasonal atmosphere-ocean forecasts. The main features of the oceanic observing systems are presented along with quality control statistics, examples of errors that can occur and some additional problematic cases.

Keywords: Quality control, Data assimilation, Hydrographic data, Bathythermographs, Temperature profiles, Salinity profiles

1. Introduction

In preparing his ocean atlas Levitus (1982) noted that "By far the biggest problem faced in this project concerned quality control of the data". The quality control (QC) checks that he used have been developed further at various centres (eg Conkright et al 2001; Bailey et al 1994) and QC manuals are available, for example, at the Marine Environmental Data Service within the Global Temperature-Salinity Profile Project (MEDS/GTSPP) website. The various systems – often relying on human input - have been designed for different kinds of data (some for temperature only), different regions and different purposes. The automated system described here processes temperature and temperature-salinity profiles of all types, and with minor variation it is suitable for both real-time and archive data and a global or regional domain. The output is specifically targeted for ocean data assimilation systems - which combine a previous gridded estimate of the ocean state with new observations to give an updated ocean estimate, often used as initial conditions for an ocean model forecast.

Quality control is an important part of ocean data assimilation systems. If erroneous values are assimilated they can cause immediate spurious overturning and also error propagation due to the sparse distribution of oceanographic data. In 2002/2003 the Met Office quality control for oceanic temperature and salinity profiles was rewritten and a comprehensive suite of automated checks introduced. The checks applied (such as spike check) have been employed in other systems. However due to missing details in the documentation, reliance on human interaction or possible improvements many of the checks were redeveloped from scratch, with trial and improvement en route.

The principles used in building the system were that:

1. The system had to be automated to cope with the data volumes involved
2. Original, reported values should be used as long as possible (they are flagged rather than deleted if they seem erroneous; the only correction currently applied is to XBT depths)
3. Any decisions taken by the system should be traceable
4. The system is designed to support data assimilation
5. Tools to monitor system performance and individual cases are available
6. The generic checks and processing use code shared with our atmospheric QC
7. The generic checks have a clear theoretical basis in probability theory

Section 2 briefly reviews the main observation types processed and their distribution and section 3 the processing and checks which are specific to these observation types. These checks on the internal consistency of each report (or sequences of reports in the track check) give a pass, fail or suspect flag to each attribute checked. These are followed by more generic background and buddy checks which are based on Bayesian probability theory (Section 4) and then a final check for vertical consistency of the flags. Section 5 gives a sample of results from the QC system and section 6 provides a summary and notes of future work.

1.1 Applications

System	Purpose	Time window	Background used
GloSea	Seasonal forecasts	7 days	Climatology
ENACT and ENSEMBLES	QC for seasonal hindcasts	1 month	Damped persistence of analysis anomalies (see section 4.6)
FOAM	Short-range forecasts	1 day	1 day forecast

Table 1. Summary of the systems which have used versions of the QC system.

Three separate applications of the QC system are summarised in Table 1. The grid used for both GloSea (Global Seasonal) and ENACT/ENSEMBLES QC systems is 1.25° in latitude and longitude, reducing to 0.3°

in latitude near the equator, with 40 levels in the vertical. The Forecasting Ocean Assimilation Model (FOAM) system includes a 20 level, 1° global model (Bell et al, 2000) and a 1/3° North Atlantic model as well as higher resolution regional models. Both FOAM and GloSea are real-time operational systems and use observations from the Global Telecommunications System (GTS) – which have had little or no QC performed at source.

ENACT (Enhanced ocean data Assimilation and Climate predicTion) was a European Commission program which aimed to A) improve and extend ocean data assimilation systems, and apply them to produce global ocean analyses over a multi-decadal period and B) quantify the benefits of the enhanced assimilation systems through retrospective seasonal climate forecasts and through analysis of ocean behaviour. ENSEMBLES is a broader European Commission program further developing earth systems models and producing estimates of uncertainty in future climate on seasonal to decadal and longer timescales.

The primary data source for ENACT/ENSEMBLES is the World Ocean Database 2001 (WOD01; Conkright et al, 2001), for dates since 1990 it is supplemented using data from World Ocean Circulation Experiment (WOCE), Australian ships (BMRC/CSIRO XBT[1] data from 1997), Pacific Marine Environmental Laboratory (PMEL; CTD reports in the equatorial Pacific described by Johnson et al, 2002) and GTSPP. Some of this 'extra' data was already in WOD01 so a duplicate check was needed. Much of the data had already been quality controlled to various separate standards, but for consistency all input QC flags were ignored[2].

The main text of this paper documents the ENSEMBLES system, differences from the ENACT system are summarised in Appendix A. The operational systems, FOAM and GloSea, currently use variants of the ENACT QC system (with different backgrounds as noted in Table 1) – they will be upgraded over the coming months.

1.2 Quality control system overview

The main steps are in order:
a) Read observations and thin vertically if necessary.
b) Correction for XBT fall rates
c) Track check
d) Constant value, spike and step checks
e) Superobbing (averaging) of moored buoy reports
f) Convert reported temperatures to potential temperature
g) Stability check for profiles that include salinity
h) Duplicate check (check for similar position/time)
i) Loose background check on reported values - Bayesian check using generic code
j) Average values that have passed QC on to model levels
k) Background check the model level values
l) Perform buddy check
m) Final multi-level check
n) Listings and statistics.
o) Output observations.

The initial checks and processing only use reported values, and up to (e) temperature is the "in situ" temperature. Subsequent checks use potential temperature, θ, (the conversion uses the reported salinity if available and not obviously wrong; otherwise the salinity from the background field is used).

---

[1] See section 2 for a description of XBT and CTD data.

[2] Our first priority had to be a QC system which did not rely on input quality flags, because some of the historical data and all of the real-time data do not have them. Intelligent use of originators' data flags, while desirable, would have required extensive preliminary investigations which we did not have resources to do.

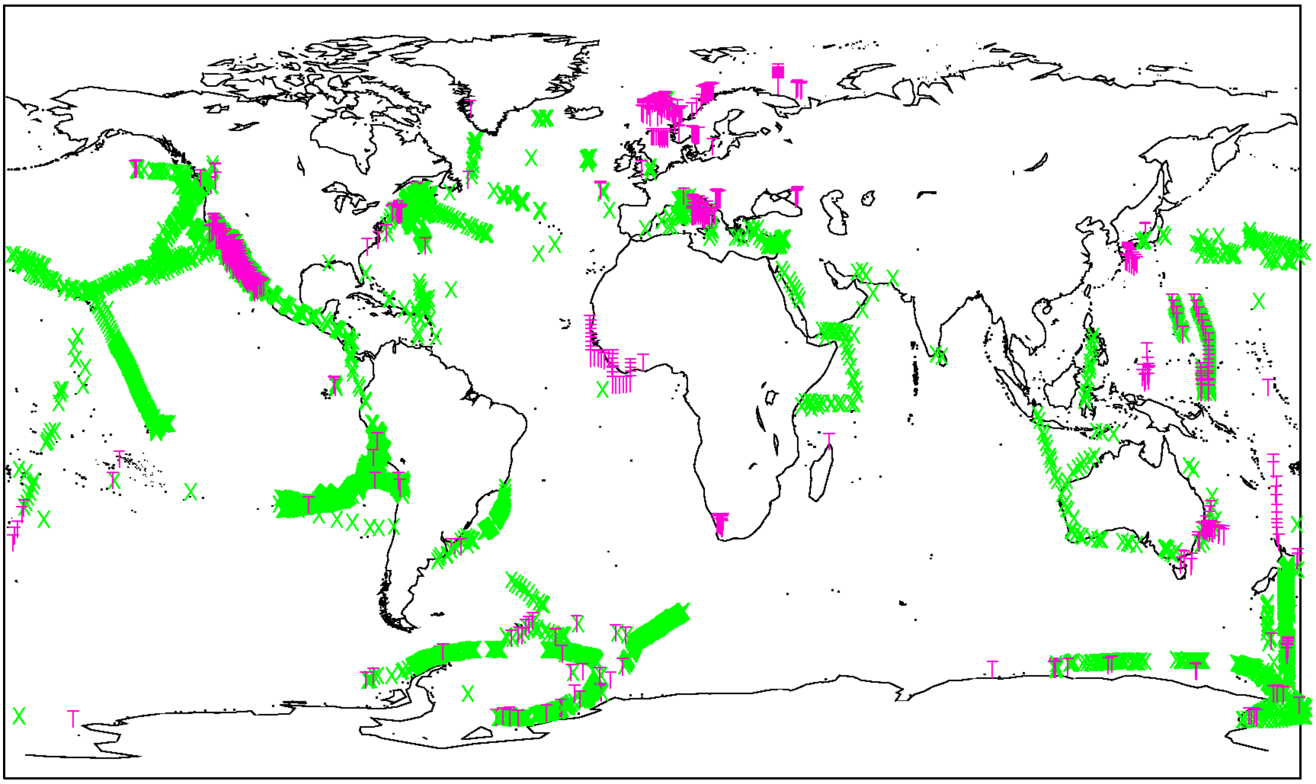

Figure 1. Reports available from WOD01 for January 1958. Green X – Bathy data (3508 reports), Purple T – Ocean station data (855 reports). Actual position is at the bottom left corner of each letter.

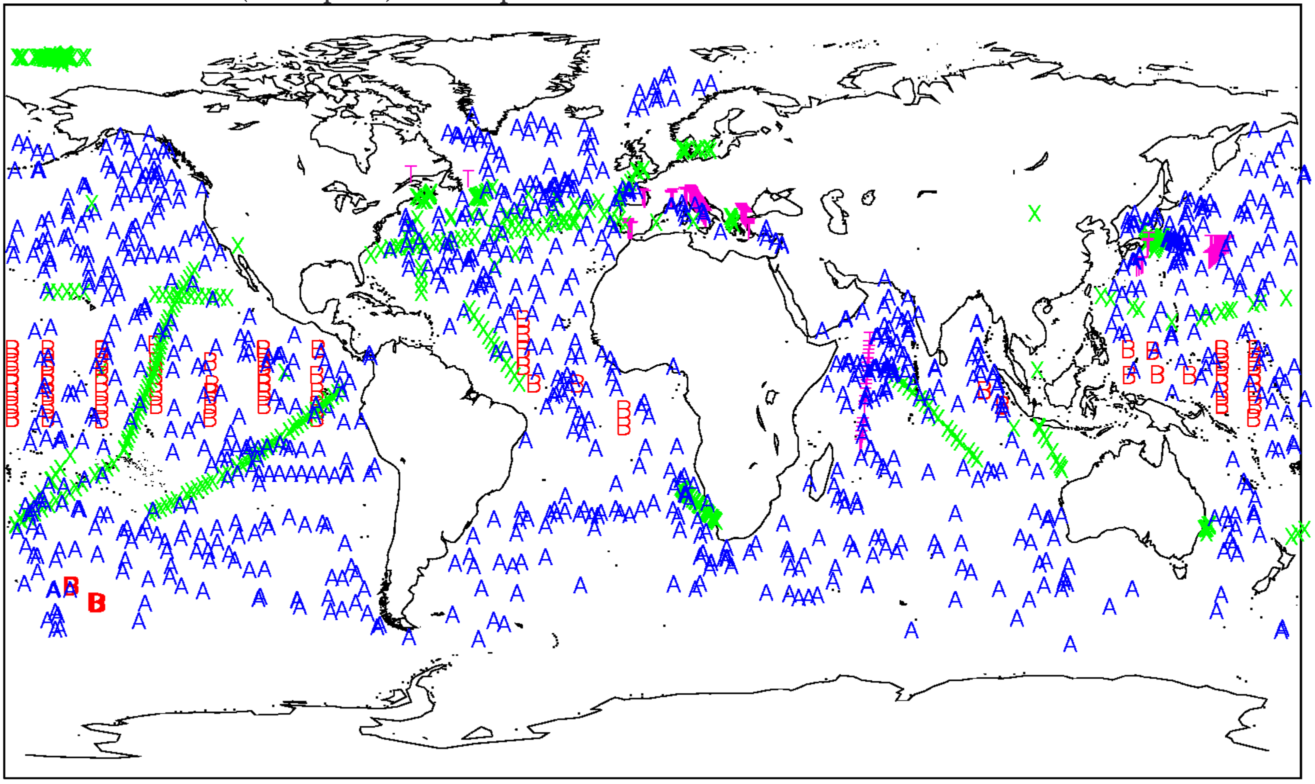

Figure 2. Reports available over the GTS for 13 to 20 September 2004 and used in GloSea. Green X- XBT data (628 reports), Purple T – CTD (217), Red B – Buoy (2899, see section 3.4), Blue A – Argo (919). There is a clear XBT position error – a report well inland in East Asia.

2. Observations processed

2.1 Bathythermographs (MBTs and XBTs)

Mechanical Bathythermographs (MBTs) consisted of a temperature recorder lowered to a depth of about 300 m and then winched up again, temperature values were read off the trace and recorded. Since the advent of expendable bathythermographs (XBTs) in the mid-1960s MBTs have gradually been phased out. Many XBTs go to either 460 m (T-4 and T-6) or 760 m (T-7) in depth, T-5 XBTs extend to about 1800 m. The temperature errors of these instruments are generally considered to be 0.1-0.2℃, reports transmitted over the GTS use the BATHY code format (with temperatures reported to 0.1℃). The depth estimation can lead to larger errors; correction of a systematic depth error is described in section 3.1. Heinmiller et al (1983) compared XBT and CTD temperatures from ten cruises (looking at 25m or depths with little vertical gradient to minimize the effect of any depth errors) and found that the samples examined were 0.19℃ and 0.13℃ warm for T-4 and T-7 XBTs respectively. Roemmich and Cornuelle (1987) and Budeus and Krause (1993) performed tank calibrations of XBTs and found negligible and 0.07℃ bias respectively. All the authors found significant probe to probe, or cruise to cruise, variability in the temperature offset.

Further details of XBTs can be found in Emery and Thomson (2001) or McPhaden et al (1998, Appendix B4; they note that about 15% of XBTs suffer malfunctions before reaching 250 m). XBTs are the most error-prone oceanic observing system; however for more than 30 years XBTs have provided much of the sampling of the ocean subsurface so good quality control of these reports is essential. Two XBT problems that we do not explicitly test for are documented by Kizu and Hanawa (2002a,b): 1) temperatures in the top few metres are affected by transient response, 2) occasionally XBT reports suffer from a temperature drift; an error typically increasing by 0.1°C per 100 m, this may start at the surface or lower down.
In WOD01 there are also Autonomous Pinniped Bathythermograph reports from instruments attached to elephant seals. These are rather localized in time and space but can have very dense coverage when present. They are processed in the same way as MBT and XBT data.

2.2 Hydrographic profiles (CTDs and predecessors)

Smaller numbers of high quality profiles are available from research vessels. These usually include salinity and sometimes other chemical concentrations. Salinity is now recorded according to the practical salinity scale defined as conductivity ratio with no unit. Numerically, the differences from concentration in parts per thousand used in earlier years are less than 0.01 within the oceanic ranges (Lewis and Perkin, 1981). In the earlier years the profiles are generally ‘bottle casts’ in which measurements are only available at about 10 or 20 levels. These were replaced by continuously recording electronic instruments, which improved over time. Emery and Thomson (2001, p39) suggest that modern Conductivity-Temperature-Depth (CTD) profiles “are accurate to approximately 0.002° C in temperature, 0.005 psu in salinity and <0.5% of full-scale pressure in depth". If transmitted via the GTS these are coded in TESAC format which uses two decimal places for both temperature and salinity.
A few of the CTD reports stored in WOD01 can have vertical resolutions of one metre and 1000 or more levels. Some **vertical thinning** is done of reports with more than 150 levels: levels are selected every 100m below 1000m, every 50m between 500m and 1000m and all the remaining available levels are divided equally in top 500m (usually giving a spacing of 5m or less).

2.3 Moored buoys

Starting in about 1989 temperature sensors were deployed on the deep ocean moorings of the Tropical Atmosphere Ocean (TAO) array in the equatorial Pacific Ocean (Hayes et al, 1991; McPhaden et al, 1998). The sensors are located at various levels down to 500 or 750 m. The vertical resolution is fairly coarse, particularly below 300 m, with only 13 or so levels in total. In 2000 the west Pacific section of the array was passed from US to Japanese operation and renamed TRITON – see section 3.4 on superobbing. In 1997/98 a smaller array of deep water buoys (PIRATA) was set up in the tropical Atlantic. There are also a number of buoys moored in fairly shallow water around North America and Europe (these often report in BATHY format on the GTS, so in our statistics they are combined with the XBTs).

2.4 Profiling floats – Argo data

During the mid-1990s PALACE (Profiling Autonomous Lagrangian Circulation Explorer) floats started to be deployed. Numbers increased rapidly after 2000 co-ordinated by the International Argo Project which aims to have 3000 floats active with a quasi-uniform global coverage. As described in Davis et al (2001) these drift at their 'parking depth' (usually 1000 or 2000 m) for seven to ten days, then rise to the surface measuring temperature and salinity and transmit the data via satellite when they reach the surface. Early deployments (some of which didn't measure salinity) were mainly in the Northern hemisphere but coverage now is more uniform. Broadly speaking their accuracy is similar to standard CTD instruments. However for salinity the unattended operation for up to four years can cause gradual drifts in salinity calibration, which it may be possible to correct using comparisons with nearby CTD data (Bacon et al, 2001). Some salinity sensors also show abrupt jumps of 0.1 or more due to biofouling (Davis et al, 2001). As yet there are no tests for such drifts or jumps in our QC system.

2.5 Data coverage

Figures 1 and 2 show global data coverage for January 1958 (from ENACT/ENSEMBLES) and a week in September 2004 (from GloSea). In January 1958 the numbers of reports are respectable but the data coverage is highly non-uniform with huge gaps in parts of the tropics and southern hemisphere contrasted with a dozen or so highly sampled ship tracks (mostly MBTs). Six months later (not shown) the total number of reports is larger but they are almost all north of 15°N. Over the following three decades data coverage improved somewhat, but was still concentrated on the main shipping lanes – with fewer reports in winter - supplemented by relatively infrequent research cruises. Figure 2 shows recent coverage including the tropical deep-water moored buoy programmes and the near-global spread of the Argo profiles. The reports are those available over the GTS within about 10 days of validity time (extra data from research cruises may be available in delayed mode).

3. Processing and checks specific to ocean data types

3.1 XBT depth correction

XBT depth is calculated from the time since release of the probe. Many XBTs fall slightly faster than suggested by the manufacturers' original equations. Based on comparisons with CTD data Hanawa et al (1995) suggest the linear correction Zcor = 1.0336 Zrep for T-4, T-6 and T-7 XBTs. Even after correction 17.5% of reports still fall outside the required accuracy of 5 m or 2% of depth. The Hanawa et al fall-rate equation was implemented by the manufacturers (TSK and Sippican) in 1995/96 and the instrument code reported was modified (although in 2005 there are still small numbers of XBTs which, according to their instrument code, are using the original fall-rate equation). A few reports prior to this were corrected by the data originators. Particularly before this change the instrument type of many XBT profiles is not known - but of those known T-4, T-6 and T-7 make up about 95%. The solution adopted for ENSEMBLES (see Appendix A regarding ENACT) is that we correct:
a) T-4, T-6 and T-7 XBTs identified as uncorrected
b) any XBTs of unknown type, reporting to less than 840 m, prior to the end of 1996

Another complication is that colder water has higher viscosity - reducing the probe fall rate. Most of the comparisons used by Hanawa et al (1995) were at relatively low latitudes. Their North-East Atlantic sample supports the suggestion of Thadathil et al (2002) (who examined CTD-XBT pairs in the Southern Ocean) that the correction needed at high latitudes is less than that at low latitudes. To model their results in a rather simple way we calculate the mean temperature of the report over the top 300 m (Tmean). If Tmean is above 15°C then a correction factor of 1.0336 is applied, if Tmean is below 5°C the factor is 1.0, with a linear transition in between. This temperature dependence is only applied when a report is being corrected anyway; arguably recent high-latitude reports should have their depths reduced.

Kizu et al (2005a,b) have recently reported that T-5 (deeper) XBTs from the TSK manufacturer have a fall rate error, and their suggested correction (Kizu et al, 2005b, table 2) has been incorporated. The number of reports affected (243 in WOD01 from 1996 onwards) is much smaller than those above.

3.2 Constant value, spike and step checks

Constant value check: If 90% or more of the temperature levels, covering at least 100 metres, are set to exactly the same value then all of the temperature values are rejected. A few realistic, fairly shallow, high-latitude profiles are rejected, ten times as many spurious profiles are rejected. (MBT data are not checked as they are shallow and low precision.) Salinity is checked for 70% of levels over at least 50 metres exactly equal – this could perhaps be relaxed to match the temperature criterion.
Spike and step check: See Appendix B.

3.3 Track check

For each ship/buoy identifier a time-sorted list of reports is set up. If the implied speed between two reports is excessive then one of them is wrong. The difficult part is deciding which position is wrong, and having an extended sequence of reports helps. A series of tests is applied: looking for excessive speeds compared to neighbouring reports, looking for a distinct kink in the track and looking for a smooth distance/time relationship in approximately collinear tracks. Reports at inland positions will be rejected as not having background values. The **MaxSpeed** is set to 15 m/s for ships and to 2 m/s for buoys - both moored and profiling. Details of the track check are given in Appendix C.

Exclusions. Some tracks can't currently be checked in a satisfactory manner – these include air-dropped XBTs and occasions when two or more ships have the same identifier. Certain default identifiers ('SHIP', '0' and ' ') are not checked. Also if 2 (NumShort+NumFast) + NumBends >= NumObs - 1
then the track is not checked. Here, NumObs is the total number of reports from the identifier, NumShort the number of intervals less than one hour, NumFast the number of speeds greater than MaxSpeed and NumBends the number of bends of 90° or more.

For oceanographic reports from the GTS roughly 1% of reports are flagged. For reports from WOD01 the proportion tends to be higher, partly because the WOD01 cruise number is not always unique to a particular ship. An example showing an exceptional number of track errors is shown in Figure 3 – each of the rejected reports is labelled with a letter. From one ship A, B and C are a 12 hour time error, a longitude error (near the coast of Brazil) and the wrong sign of latitude. O, P and L have erroneous longitudes. Most of the other rejections are due to two or more reports at the same nominal date/time but with different positions – presumably because a batch of reports has been compiled or transmitted together – this seems to be more prevalent in WOD01 than GTS data. The east-west track near G has a kink, probably from an erroneous latitude, but this was not detected as it was not a large enough error to give an excessive speed. Of course, what constitutes an error depends on how time errors are regarded and at what distance spatial errors become significant. Figure 3 shows results from ENACT, for ENSEMBLES the track check was relaxed in some respects. In ENSEMBLES reports G to N and P were not flagged – beneficial in most cases, but L and P have real if modest position errors.

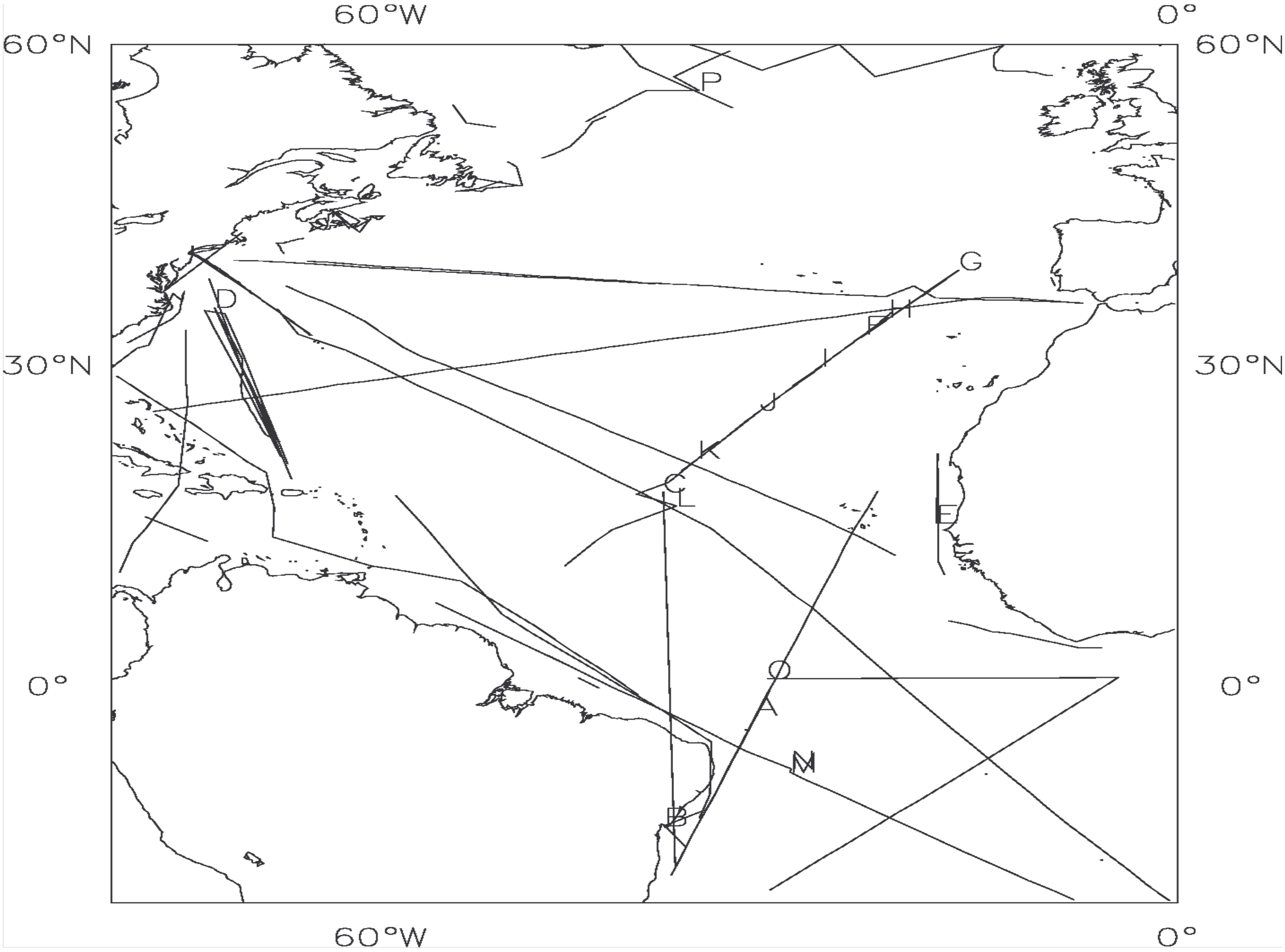

Figure 3. Ship tracks and detected track errors (denoted by letters) for part of the Atlantic, October 1996, ENACT QC system.

3.4 Superobbing

In the tropical Pacific the TAO buoys report daily averages. The TRITON profiles are quasi-hourly and are sometimes split into two GTS reports with different sets of levels. For consistency in data assimilation the TRITON data are formed into daily averages. For each level a simple average is calculated (duplicate and rejected values are not used). The standard deviation (SD) of the reported values at each level is also calculated and if this is excessive then the averaged value is omitted. For salinity a threshold of 0.5 is used, for temperature it is 0.7/2.5/2.0/0.8/0.4 °C for levels of 0-40/41-200/201-250/251-500/501+ m respectively: allowing for large SDs within the thermocline (partly due to the semi-diurnal tide). In figure 2, 2431 of the 2899 buoy reports were hourly data, the daily averaging produced 110 superobs from these, giving a total of 578 Buoy reports to be assimilated.

3.5 Stability check

Reports with both temperature and salinity are checked for density inversions. The McDougall et al (2003) equation of state is used to calculate density (ρ) as a function of θ, salinity (S) and pressure (P). When comparing levels *k-1* and *k*, we compare the (potential) density at level *k*

$$D\rho_k = \rho(\theta_k, S_k, P_k) - \rho(\theta_{k-1}, S_{k-1}, P_k)$$

we allow small inversions and test for $D\rho_k < -0.03\ kgm^{-3}$. There is a test for "spikes" in the density:

$$\left|D\rho_{k-1} + D\rho_k\right| < 0.25\left|D\rho_{k-1} - D\rho_k\right|$$

in this case θ and S at level *k-1* are flagged, otherwise $D\rho_k$ and $D\rho_{k+1}$ are tested in the same way (and level *k* flagged if the test is true). If neither of these tests is positive it is not clear whether level *k-1* or *k* is erroneous and temperature and salinity at both these levels are flagged as suspect. If there is a significant inversion at the bottom of the profile then the lowest values of temperature and salinity are flagged (errors in the level above are possible but less likely). If the number of inversions in a report is ≥ MAX(2,NumLev/4), where NumLev is the number of levels, then the whole report is rejected. In general this affects only a rather small proportion of reports, but in the 1960s and 1970s there were higher stability rejection rates in the boreal winter particularly in the seas around Japan.

3.6 Duplicate check/thinning

The duplicate check looks for pairs of reports within 0.2° latitude/longitude and 1 hour, the report with the highest preference factor is retained (one is chosen at random if they have the same preference factor). The preference factor is set to the number of levels in the report, plus the penultimate report depth (in 100s of metres), plus 100 if the report is not a Bathy, plus 10 if the identifier is not 'SHIP', ‘0’ or blank, plus 20 for WOD01 data. Thus there is a preference for deeper reports, with more levels and salinity if possible. In some areas dense sequences of XBT reports are thinned by this check – for some data assimilation applications this is desirable.

3.7 Prior rejects

XBT temperatures below 1000 m are rejected. Their root mean square differences from background are significantly worse than those for CTD data whereas above this level they are fairly similar (section 5.1) and about half of the deep XBT profiles examined had quality problems below 1000 m.

A limited examination of analysis increments (section 4.6) revealed two cruises with particularly poor quality temperature profiles and another cruise with a bias of about 0.2 in salinity; these were rejected. Some of the poor temperatures were in western boundary currents where the QC is less likely to flag the data because of larger error estimates.

4. Background and buddy checks

We follow Lorenc and Hammon (1988) and Ingleby and Lorenc (1993) in using Bayesian probability theory to provide a theoretical basis and to combine information from different checks. They suggest a model of observation errors, which will be summarised here. Some observations have “gross errors” from communication errors, instrument failure etc, these are assumed to have a flat probability distribution with density κ (roughly the reciprocal of the climatological range of the variable) – they give no information on the true value. “Good” observations, without gross errors, have errors that are assumed to be normally distributed with standard deviation σ_o and mean zero. Each observation value is assumed to have a climatological or *a priori* Probability of Gross Error (PGE); the background and buddy checks update the PGE for each value and values with $PGE_{final} \geq 0.5$ are rejected. The specified parameters are:

a) κ is set to 0.1 for temperature and 0.25 for salinity (corresponding to error ranges of 10°C and 4)
b) the initial PGE is set to 0.01 (ie assuming that 1% of reports are erroneous) except that it is set to 0.05 for bathythermograph temperatures
c) Values considered as suspect by the spike/step check or the stability check have their initial PGEs

increased: PGE = 0.5 + 0.5*PGEclim[3]
Note that the background check is more sensitive to the specified σ_o and σ_b (see below) than it is to these parameters.

4.1 Observation and background errors

The ocean has a lot of variability on scales smaller than the ocean model grids (usually 100 km or more) currently used for global ocean forecasting. For data assimilation purposes this variability sampled by the point observations but unresolved in the model is unwanted noise and is referred to as representivity error.

observation error = measurement error + representivity error

For these ocean models the representivity error is significantly larger than the measurement errors (between about 0.002°C and 0.2°C for temperature – section 2) so that data assimilation systems use estimates of σ_o for temperature between 0.5°C and 1°C (Wang et al, 2002, and Weaver et al, 2003). Salinity assimilation is a fairly recent development and associated error estimates are rather lacking in the literature.

For our system σ_o has been estimated from ENACT and preliminary ENSEMBLES results. For temperature σ_o is 0.78°C at the surface, rising to 1°C at 75 m, then reducing to 0.07°C at 4500 m and below. For salinity σ_o decreases from 0.18 at the surface to 0.013 at 5500 m. More detail is given in Table 3 later.

The background error SD σ_b values vary geographically and will (with exceptions, noted below) be similar to climatological SDs – with maxima in the area of the Gulf Stream for example. The World Ocean Atlas 2001 (http://www.nodc.noaa.gov/OC5/WOA01/) 1°×1° SDs for temperature and salinity were averaged onto a 2°×2° latitude-longitude grid and minimum (set as σ_o) and maximum values set level by level. The fields were then smoothed horizontally and filled in by interpolation where necessary. Both observation and background temperature error estimates have a maximum round the thermocline (as in Weaver et al, 2003).

Between 14°S and 14°N a maximum temperature error estimate of 1.7°C is used – this represents the fact that the persistence background used has significantly lower errors than a mean climatology during El Niño events. In the Black, Caspian and Aral Seas temperature error estimates were set to be at least $4\sigma_o$. In the Arctic (North of 70°N and down to 400 m) salinity errors are set to be at least $2\sigma_o$ - the factor tapers off to the South and vertically. In the Baltic Sea salinity errors are set to be at least $4\sigma_o$ - the factor tapers off across the North Sea. However the Baltic, and to a lesser extent the North Sea, does have large temporal variations in salinity. This (perhaps combined with problems in the initial conditions/climatology) means that despite the enhanced salinity error estimates, there are still excessive salinity rejections in this area.

Information about climatological covariance scales in the tropical Pacific is given in Kessler et al (1996), these show pronounced East-West stretching near the equator, see also Stammer (1997). Atmospheric fields show a related but less marked East-West stretching in the tropics, and forecast error scales are generally shorter than climatological covariance scales (see Ingleby (2001) and references). The latter characteristic is not modelled - forecast and climatological backgrounds are given the same correlation scales in the buddy check.

4.2 Background check

The background is 1-day ocean forecast for FOAM, a climatological field for GloSea or a damped anomaly persistence forecast for ENACT/ENSEMBLES (section 4.6) - these will have different errors. The background values are interpolated horizontally to the latitude/longitude of each observation. Bilinear interpolation is

---

[3] In principle the various checks in section 3 could have been cast in probabilistic form, but for simplicity this was not done. It would be difficult to estimate eg the probability of two position errors in a particular ship track and there can be excessive computational demands in considering all possible combinations of errors.

used if possible, if not then the background value at the closest valid grid point is used. If all surrounding grid points are missing then the background value is set as missing and the observed value will fail the background check. The profile on model levels is then vertically interpolated (linear in depth) to reported levels as necessary. Background error estimates are interpolated in the same way.
The Bayesian background check is as described by Lorenc and Hammon (1988). The PGE given the background value is

$$P(G\,|\,O) = \kappa P(G) \big/ \left(\kappa P(G) + (2\pi V)^{-0.5} \exp(-(o-b)^2 / 2V)(1-P(G))\right) \qquad (1)$$

Where *P(G)* is the prior PGE and $V = \sigma_o^2 + \sigma_b^2$. (Within 10° of the equator the temperature σ_b is multiplied by a factor of 1.5 if a climatological temperature background is used.) Because V decreases below the thermocline the rejection threshold generally decreases with depth. The background check is applied to the reported level values – with σ_b multiplied by 1.5 (2.0 for a climatological background) to make the check less strict. After vertical averaging the background check is reapplied to the model level values and the output probabilities form the input to the buddy check.

4.3 Vertical averaging

Vertical averaging onto model levels is needed for the buddy check and also acts as a preprocessing step for the Met Office assimilation system. Currently each unflagged reported level increment is assigned to the nearest model level and a simple averaging performed. For ENACT/ENSEMBLES a later step maps back flags from the model level value to the reported level value(s) that contributed to it. Sharp changes in vertical gradient can cause problems for averaging or interpolation methods, and where there is a sharp, tropical thermocline, shallower in the background than in the report, the averaging of observation minus background (*o-b*) increments can result in a small warm 'nose' at the bottom of the mixed layer in the averaged profile.

4.4 Buddy checking

The averaged (model level) data are buddy checked using as input the PGEs from the background check. Pairs of close profiles are found and the *o-b* increments at the same depth (model level) are compared and the PGEs updated. The comparison and updating are performed separately for temperature and salinity. At the end of the buddy check those values with PGEs ≥ 0.5 are rejected. Note that the buddy check can either increase or decrease the PGE. The buddy check is essentially as developed by Lorenc and Hammon (1988) and reformulated in Ingleby and Lorenc (1993, a ‘damping’ factor of 0.5 is used as described in their Appendix C). Following Martin et al (2002) the background error covariance is modelled by two SOAR (Second Order AutoRegressive) functions with length scales $l_{mes} = 100\ km\ and\ l_{syn} = 400\ km$. Within 10° of the equator the East-West scales are doubled but the North-South scales are unchanged, the anisotropy factor decreases linearly to unity at 15° latitude. There is a Gaussian correlation in time with $T = 432000\ \sec = 5\ days$. Thus the covariance is

$$C = \sigma_{mes,a}\sigma_{mes,b}(1 + r/l_{mes})\exp(-r/l_{mes} - t^2/T^2) + \sigma_{syn,a}\sigma_{syn,b}(1 + r/l_{syn})\exp(-r/l_{syn} - t^2/T^2) \qquad (2)$$

where $\sigma_{mes,a}$ is the mesoscale error variance at observation *a* etc, *r* is the horizontal distance and *t* the time difference between the two observations.
Profiles within 400 km are allowed to buddy check each other (in rare cases this may result in inappropriate pairings). Buddies with the same identifier are not allowed - testing allowing such pairs showed some problems with biased observations (ie correlated errors). In data sparse areas (much of the ocean until recently) the buddy check will have very limited effect.

Error correlations are thought to be higher along density surfaces than at constant depth; this could be taken account of in the choice of which levels to compare from different profiles but there would be complications. Ideally the correlations would also reflect differences in bottom topography and any barrier (a peninsular or isthmus) between any two points. However modelling such effects (particularly in the analysis - where it would be more important than in the buddy check) is far from trivial and will not be attempted for now.

4.5 Final multi-level check

It was found that, because of the level by level nature of the background check, that a final check on vertical consistency of the QC decisions was required. If more than 50% of the temperature levels have been flagged then the whole report is rejected. If more than 50% of the salinity levels have been flagged then all the salinities are rejected. The percentage flagged is calculated from the sum of the percentages for reported levels and averaged levels - because flagged reported levels don't contribute to the averaged values. There is then a step to reinstate flagged T or S values if they agree with values above/below - typically these will be values with slightly larger *o-b* differences, sometimes in the thermocline. A tolerance is defined: 0.5 °C, or 0.1 pss down to 200 m, changing linearly to 0.25°C, or 0.05 pss below 300 m (the transition depths are 100 m deeper within 20° of the equator). If a flagged model level value matches an adjacent unflagged value to within the tolerance, or if the values above and below are both unflagged and the value lies between them, then it is reinstated.

4.6 Objective analyses and background fields for QC

In the quality control of atmospheric data it has long been common practice to use a short-range forecast in the background check. In the quality control of oceanographic data it has been fairly common to use a check against climatology instead. In general terms ocean forecasts are somewhat better than climatology near the surface, but they may be worse at depth due to model drift and the sparsity of observations to correct the model. For the ENACT/ENSEMBLES quality control system we implemented a system which, as far as possible, combines the best features of methods without the computational expense of a forecast model. Each month a 3D analysis of (potential) temperature and salinity is performed using the Analysis Correction scheme as described by Bell et al (2000) but updated to use two horizontal SOAR functions (with scales of 300 and 400 km – but stretched E-W near the equator). The background for the following month is then given by $B_{i+1} = C_{i+1} + \alpha(A_i - C_i)$ where C_i is the climatology (averaged over all years) for month i, and A_i is the analysis and α is a factor between 0 and 1. Thus we are using a damped persistence forecast which relaxes to climatology in the absence of recent observations. This is similar to a system used by Smith (1995), and like him we find particular benefit in the upper layers of the tropical Pacific which have a large interannual variation. After experiments with various different values of α we used 0.9.

Figure 4 shows that during an El Niño/ La Niña event the thermal structure of the equatorial Pacific has large deviations from climatology (for temperature the deviations can be several times typical values of σ_b). The climatology used above, and in figure 4 for comparison, is the World Ocean Atlas 1998 (WOA98, available from www.nodc.gov). The solid line is derived from the TAO/PMEL gridded analysis of the 20°C isotherm depth (McPhaden, pers. comm.). The ENACT QC background captures most of the variability but with a time lag of one or two months (this is with $\alpha = 0.9$, using smaller values of α gives a larger time lag and extrema smoothed a bit more) and the objective analysis matches the TAO/PMEL product quite closely, with a time lag of about half a month. Use of this background gives slightly lower QC rejections than use of climatology. To initialise the system the background was set to the WOA98 climatology. The monthly objective analyses are a valuable by-product of this processing - they are independent of any forecast model, they have been used in the validation phase of ENACT and are available for other researchers (see http://www.ecmwf.int/research/EU_projects/ENACT/index.html). Clearly the quality of the objective analyses will vary spatially according to the availability of data and also depend on the error covariances used.

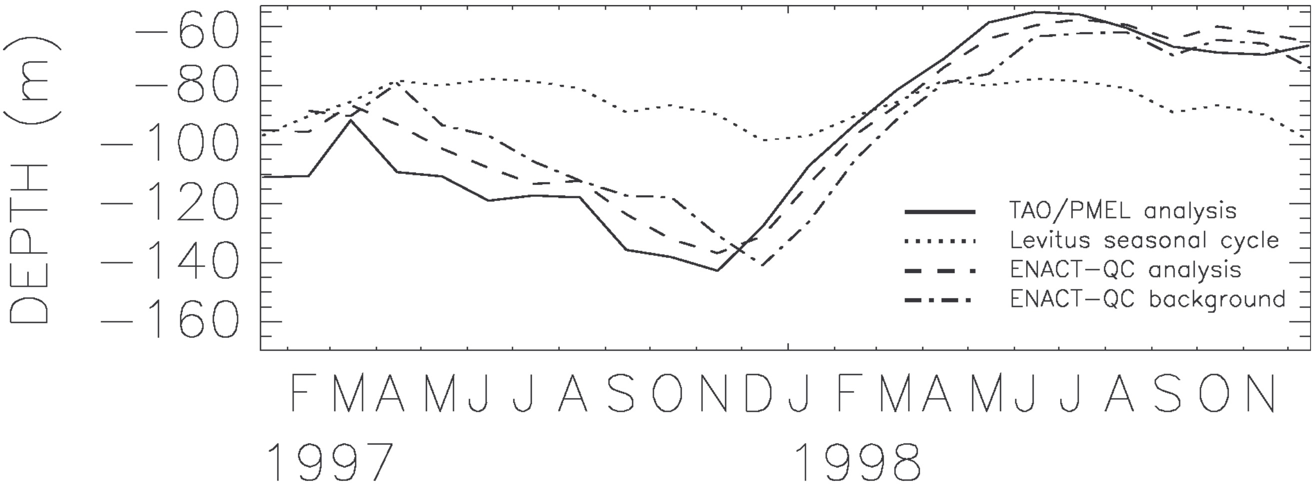

Figure 4. Depth of 20°C isotherm over the NINO3 region (150°W - 90°W, 5°S - 5°N) ENACT QC background and analysis using $\alpha = 0.9$.

5. ENSEMBLES: QC Results

5.1 Mean and Rms statistics

Figure 5a shows temperature mean and root-mean-squared (rms) *o-b* statistics for the whole of the period processed for ENSEMBLES. The observed values are those averaged onto model levels that passed the QC and thinning process. The background values are as described in section 4.6. The categories presented are those in section 2, except that hydrographic and Argo profiles have been combined as 'Tesac' data. In the top 200 m there is significant noise for the Buoy (mainly TAO/TRITON) data. This is due to some levels having very little buoy data and different sets of measurement depths being used in the West and East tropical Pacific. The Buoy data has rms values of about 1°C near the surface, then a maximum of about 2°C (larger if a climatological background is used) between 65 and 150 m – typical thermocline depths for that area – decreasing to 0.8°C at 250 m and less below.

The Bathy and Tesac rms values are fairly similar, increasing in the top 50 m to a maximum of about 1.7°C for Bathys (about 0.2°C more at these depths for Tesacs), then gradually decreasing in close agreement down to 800 m. In the top 200 m some levels have 40% larger samples than adjacent levels because they lie close to 'standard' levels giving some noise; below 500 m the numbers of Bathy values decline sharply. The generally similar rms values for Bathy and Tesac supports the idea that representivity error tends to dominate measurement error. However there are clearly differences in geographical sampling with the mean 5 m Tesac temperature being 15°C, compared to 19°C for Bathys and 26°C for Buoys (the biggest difference between the Tesac and Bathy distribution seems to be that there are proportionately more Tesacs in the Northern North Atlantic). In terms of mean difference from background there is an off-set of about 0.1°C between Bathys and Tesacs with Bathys relatively warmer - this is fairly robust when different periods are considered, and broadly consistent with the XBT biases discussed in section 2.1.

The Tesac salinity statistics (Figure 5b) are simpler, showing a surface maximum in rms and a decrease with depth. In the top 150 m very large rms values from the Baltic inflate the overall rms and account for the spikes. The salinity sensors on moored buoys are a recent development and the sample size is relatively small.

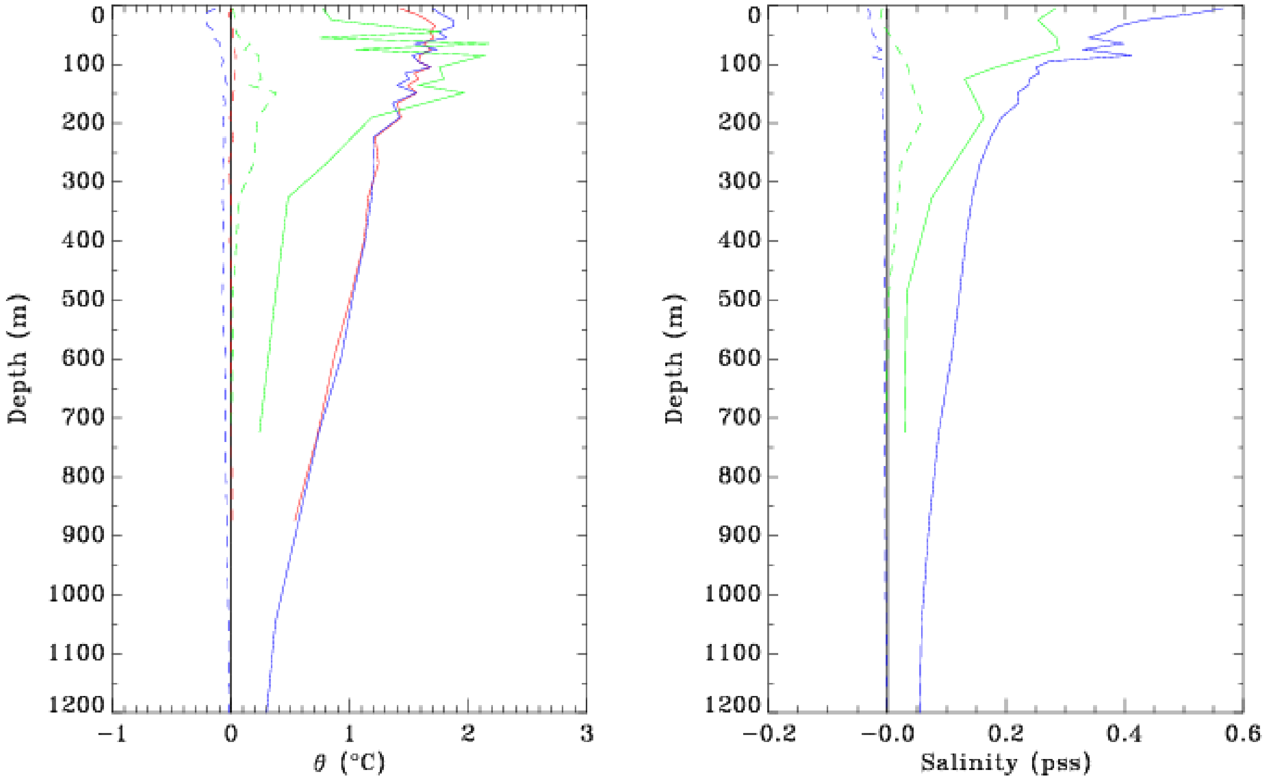

Figure 5. a) temperature and b) salinity *o-b* statistics for Tesac (blue lines), Bathy (red lines) and Buoy (green lines) data from ENSEMBLES. Mean (dashed) and rms (solid) values are shown.

A plot of Bathy rms versus time (Figure 6) does show significant development, as well as an annual cycle in the top 100 m. Until about 1967 only the relatively shallow MBT data is available, there is then an increase in the depths reached and a temporary increase in rms at some levels – due to 30% or more of XBTs being located in the Gulf Stream in the affected months. In the mid-late 1990s there was been some reduction of rms differences above 400 m. Below this level there are fewer reported values and the temporal evolution is more confused, with large rms episodes in the 1970s and early 1980s. As noted in section 3.7 XBT values below 1000 m had relatively large rms and the values were rejected in the ENSEMBLES processing. The rms Tesac statistics exhibit little year-to-year trend (not shown) for temperature, for salinity below 500 m rms values since 1990 are generally smaller than previous years, perhaps representing improved instrumentation. However both T and S show smaller rms values in the top 200 m since 2001 – perhaps due to Argo data.

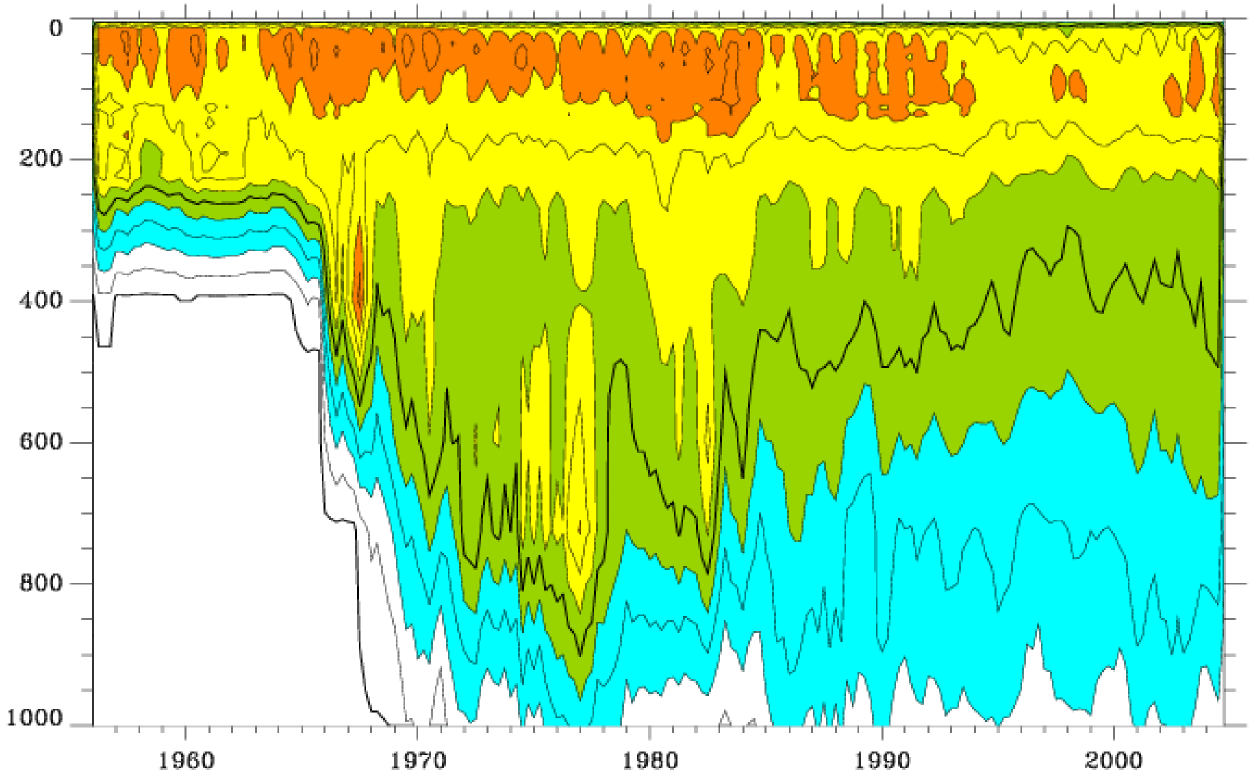

Figure 6. Temperature rms *o-b* statistics for Bathy as a function of time. Contour interval is 0.2°C, zero line (indicates less than 20 reports per quarter year) and 1.0°C line are thicker.

5.2 Numbers reported and flagged

For the 1956-2004 period the numbers of observations processed for ENSEMBLES was:
Bathy – 4.2M, Tesac – 2.3M, Buoy – 0.9M giving a total of 7.4 million reports, including duplicates.
Figure 7a) shows the total number of profiles of all types available per month and Figures 7b) and 7c) show the numbers of model level temperature and salinity values – a better measure of the number of 'independent' pieces of information (see Table 3 for an indication of the model levels). The jump in numbers in 1990 (Fig 7a) is due to duplicate reports - not reflected in the other plots. Pre 1990 about 20% of reports are removed as (near) duplicates, during the overlap between WOD01 and GTSPP it is just over 40%. Up till about 1985 there is generally an increase in the number of reports. From about 1994 to 2000 there is a decrease in the numbers of usable temperature and salinity values – particularly the latter – this is most likely due to reports from research cruises which had not been publicly released in time for WOD01. Since 2001 there is an increase particularly in the number of salinity values (Fig 7c) due to Argo data. The other obvious features are that there are three to five times as many temperature values as salinity values, and that for many years there is a clear seasonal cycle with fewer reports in the boreal winter.

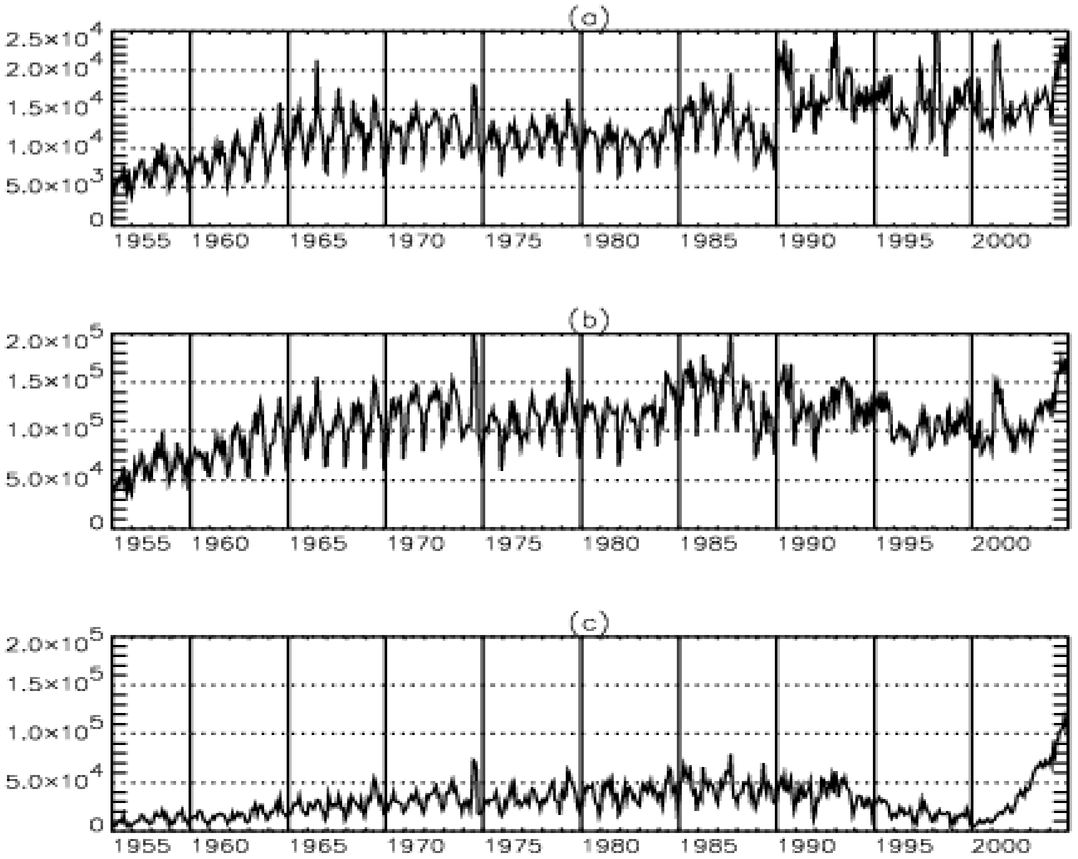

Figure 7. a) total number of profiles available per month, b) number of temperature values on model levels, c) number of salinity values on model levels

Table 2 shows the percentage rejection rates at different stages in the processing (duplicate reports are excluded). The first two columns give the rejection rates from the loose background check on reported level data and the tighter background check on model level data. The buddy check only makes modest changes to the latter figure – the final checks (Section 4.5) have more effect. The final model level flags are then copied back to the reported values within each model layer – these plus the 'backgr(rl)' flags and any from the checks described in Section 2 (with the track check as the largest contributor) give the 'overall' rejection rate.

From this the moored buoy and Argo data have the lowest temperature rejection rates. CTD+ temperatures have fewer background rejections than Bathy temperatures. Salinity rejection rates are higher, over 6% for CTD+ and 1% for Argo (the Argo salinities are almost all since 2001). There are small numbers of moored buoy salinities – with a rejection rate of about 1%. Time series (not shown) indicate that CTD+ salinity flagging gradually decreased since about 1980; for reasons that are not understood the Bathy temperature flagging rate increased in the early 1990s and has stayed higher since. Both the Buoy and Argo data had peaks in rejection rate in the first few years of deployment (but sample sizes were fairly small at that stage) then settled down to rather low rejection rates. Broadly speaking the rejection rates will reflect the proportion of 'bad' values, although we would expect the rates to decrease slightly where background errors are relatively small (in the later years of the period as the observation coverage increases, and particularly where there are moored buoy arrays). In some months there are spikes in the rejection rate: partly from observational quality problems, but also from observational campaigns in particular areas of the oceans – as previously noted the background can be relatively poor in the Arctic, Baltic and Black Seas and round the Gulf Stream and Kuroshio currents.

		backgr(rl)	backgr(ml)	final(ml)	overall(rl)
Bathy	T	0.9	0.9	1.3	3.0
CTD+	T	0.2	0.4	0.5	1.6
Buoy	T	0.2	0.1	0.1	0.3
Argo	T	0.1	0.2	0.2	0.4
CTD+	S	3.1	1.6	2.1	6.2
Argo	S	0.7	0.2	0.3	1.0

Table 2. Percentage rejection rates 1956-2004. CTD+ includes earlier “bottle” data; rl denotes reported levels, ml denotes (values averaged onto) model levels. For other details see text.

5.3 Error estimates

		ß	Temperature		à	ß	Salinity		à
level	depth	N/1000	rms	$\overline{\sigma_o}$	$\overline{\sigma_b}$	N/1000	rms	$\overline{\sigma_o}$	$\overline{\sigma_b}$
1	5	1623	1.70	0.78	2.77	1313	0.566	0.18	0.40
2	15	1380	1.77	0.80	2.63	1141	0.510	0.17	0.36
3	25	1387	1.87	0.85	2.46	1176	0.432	0.17	0.33
4	35	1154	1.88	0.90	2.22	977	0.390	0.17	0.30
6	55	1132	1.82	0.96	1.86	982	0.340	0.17	0.27
8	75	1062	1.74	1.00	1.61	921	0.330	0.17	0.25
11	105	915	1.68	0.94	1.46	752	0.251	0.17	0.22
13	125	564	1.50	0.90	1.29	463	0.239	0.16	0.21
15	149	850	1.56	0.85	1.32	732	0.220	0.15	0.20
17	190	879	1.42	0.77	1.21	759	0.192	0.14	0.18
19	268	716	1.21	0.65	0.98	636	0.157	0.12	0.15
20	326	689	1.18	0.59	0.92	613	0.143	0.11	0.13
22	488	620	1.03	0.51	0.76	568	0.122	0.09	0.11
24	725	389	0.74	0.39	0.57	356	0.087	0.06	0.07
26	1046	418	0.37	0.29	0.34	390	0.058	0.04	0.05
28	1460	243	0.22	0.18	0.21	237	0.046	0.03	0.04
30	1972	164	0.12	0.10	0.12	161	0.030	0.03	0.03
32	2582	60	0.11	0.09	0.09	60	0.021	0.02	0.02
34	3258	38	0.11	0.08	0.09	38	0.018	0.02	0.02
36	3948	32	0.09	0.08	0.08	31	0.016	0.02	0.02
39	4983	11	0.07	0.07	0.07	11	0.016	0.01	0.01

Table 3. Statistics for TESAC data (ocean stations, CTDs, Argo) for 1956-2004, selected model levels. Rms(*o-b*) values and mean error estimates at report locations. Only data that has passed QC and thinning checks is used.

If, as usually assumed, observation and background errors are uncorrelated and unbiased then we should have $(rms(o-b))^2 \approx V = \sigma_o^2 + \sigma_b^2$. Given the complex nature of background and observation errors and the simplifications made in modelling them we should not expect exact agreement, but (assuming the sample size is large and representative) *rms(o-b)* gives useful guidance for σ_o and σ_b and provides an upper bound for them. Comparison of the rms and error estimate values in table 3 suggests that the latter are approximately correct at most levels. The main exception is that in the top 50 m the temperature σ_b estimates are too large. This is mainly due to the near-surface background being closer to the time-varying reality than a climatological mean can be.

5.4 Large gradient cases

Particular problems occur in areas of large vertical and/or horizontal gradient. An example from ship tracks crossing the Gulf Stream is shown in Figure 8. There are 14 XBT profiles in a box of 1.5° latitude by 1.5° longitude, most within a three day period: they fall into warm and cold clusters. At 100 m depth the reports range between about 4°C and 17°C – below this depth the warm values decreases steadily. There is little doubt that the reports are correct in essence if not in all details (extending the area there are further profiles in both clusters). The background is too smooth to represent this gradient and lies slightly closer to the warm cluster. A few of the colder values were rejected by the QC, but most values were accepted – mainly because of the large background error estimates in this area. There are also large vertical temperature fluctuations in individual profiles, but these seem realistic for such an active area (unfortunately there are no salinity profiles in the area to confirm that there are compensating salinity fluctuations). The Kuroshio current also has large gradients, but possibly not quite as large as the Gulf Stream.

At low latitudes the thermocline tends to be rather abrupt, an example from the eastern Indian ocean is shown in Figure 9. There are five XBT profiles: four of them are high vertical resolution, one of these is a duplicate (actually one hour later than its 'twin' but very similar). The lower resolution profile has a spike at 109 m (rejected by the spike check, it would have passed the background check), at 550 m it jumps by 0.5°C but the three slightly offset values were not flagged. The sharpest thermocline here is for the two reports at 5°N – about 11°C between 100 and 150 m. The background thermocline is too smooth and somewhat too shallow thus giving *o-b* differences of up to 6°C at 130 m and down to -3°C at 160 m – these have passed the background check but there are more extreme examples which have resulted in rejected values.

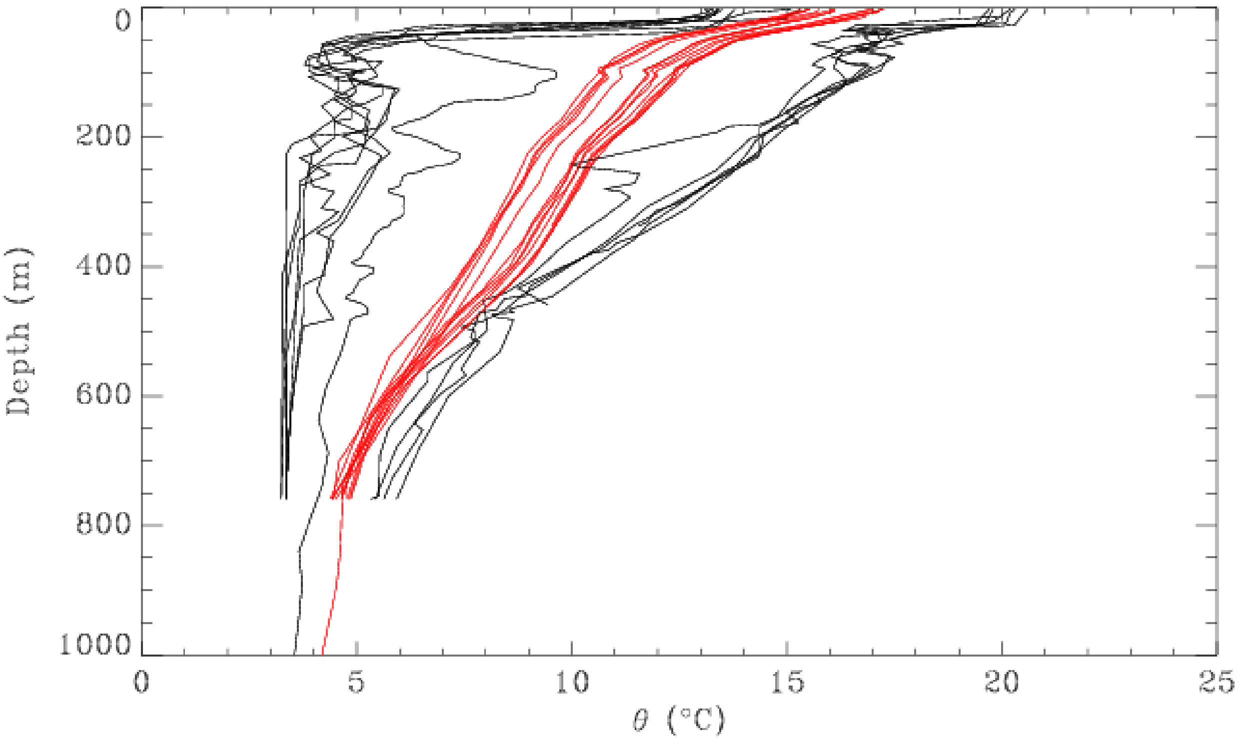

Figure 8. ENSEMBLES, July 1998, profiles within 45.5-47°N and 43.5-42°W. Black lines – reported profiles, red lines – corresponding background profiles.

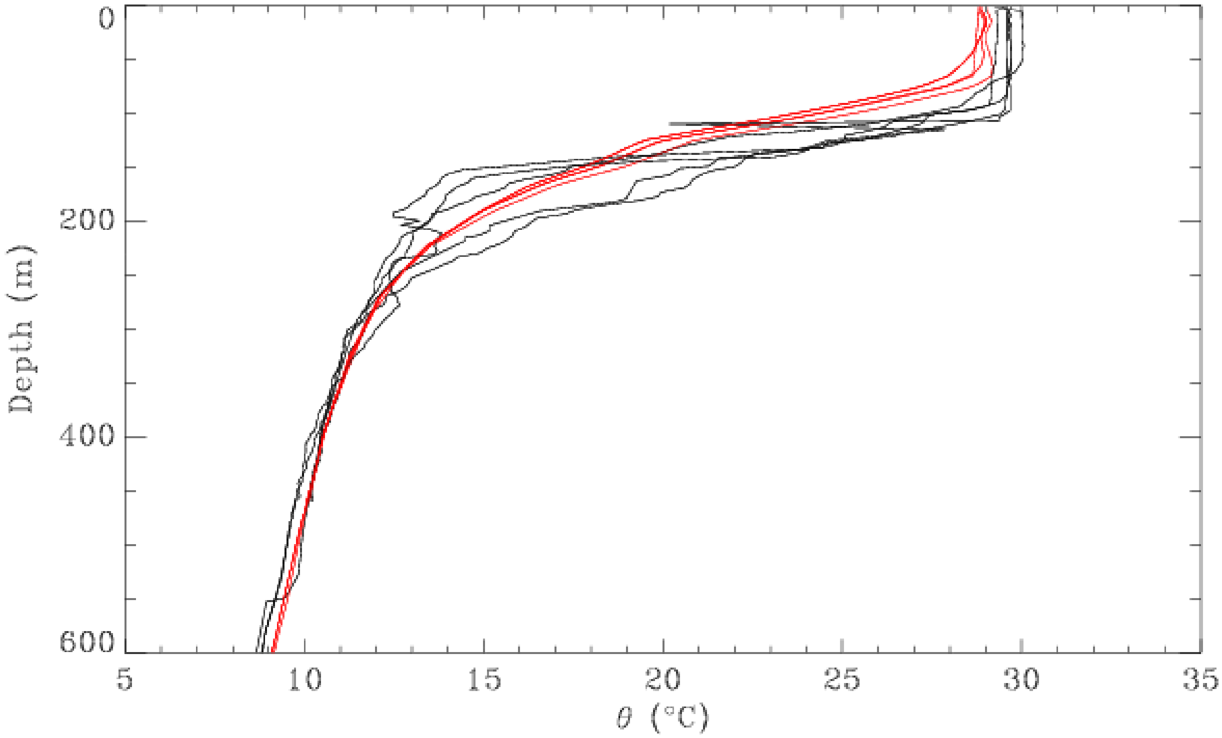

Figure 9. As figure 8 but for 5-6.5°N and 93.5-95.5°E.

6. Discussion and summary

The recently developed Met Office quality control system for temperature and salinity profiles has been described and some results from its use with archive data have been presented. It is also used in two real-time configurations. Overall it appears to be performing well although there is scope for improvement in any QC system. Metadata are important: XBT fall-rate issues meant that some reports from 1995 onwards were 'double-corrected' in ENACT – now set right; ship identifier and time problems make the ship track check more difficult. Stability and other checks could be tightened up in some geographical regions. Reports that have been flagged at some levels but not others would benefit from a more sophisticated final check for bias or drift in the reported values.

In some cases examined the decisions could have been improved with human intervention – but comprehensive intervention is highly time-consuming. In our opinion track checking and looking for systematically poor cruises would be the most effective use of any intervention. An intercomparison of ocean QC systems is planned under the auspices of the Global Ocean Data Assimilation Experiment (GODAE). However preliminary investigations suggest that intercomparison is time-consuming and QC-decisions are not always clear cut. In any QC system there is a balance between trying to reject all 'bad' observations and retain all 'good' ones – different users might require a different balance. This balance is especially challenging in areas of large gradients and small scale eddies such as the Gulf Stream.

As part of the ENACT and ENSEMBLES projects the QC system has processed 7.4 million reports (including duplicates) of various types over the period 1956-2004. The latest version of the processed data is now generally available (via www.hadobs.org). About 20% of reports were discarded by the thinning (duplicate check). Of the remainder about 0.5% of Argo and Buoy temperature values were rejected, but 2-3% of ship-based temperature values were rejected and 6% of salinity values were rejected. XBT reports are prone to several types of errors and also show a small warm bias on average.

The background check, and the background and the background error estimates used in it, are important. An improved background, along with a reduced error estimate, makes the check more discriminating. The GloSea and FOAM systems would benefit from incorporating elements of the ENACT/ENSEMBLES background calculation – using recent observations where available, but generally close to climatology in data sparse areas and at depth. The arguments and statistics presented here support the view that for current global ocean models the unresolved detail (representivity error) tends to be larger than the actual measurement error of the instruments used. The observation error estimates (including representivity error) presented here are more sophisticated than those previously used in ocean data assimilation.

Acknowledgements

The QC development and historical processing were part of the ENACT and ENSEMBLES Projects funded under EU contract numbers EVK2-CT2001-00117 and GOCE-CT-2003-505539. Clearly the provision of data archives from NODC, WOCE, BMRC, PMEL and GTSPP has been fundamental to the historical processing; Tim Boyer and Melanie Hamilton have been very helpful in understanding the data. Equally fundamental are the research and operational programmes that make the observations, including the International Argo Project and the national programmes that contribute to it (http://www.argo.ucsd.edu, http://argo.jcommops.org). Within the Met Office thanks are due to Adrian Hines, Matt Martin, Mike Davey and Mike Bell who have contributed in various ways to this work. Thierry Carval of IFREMER helped with understanding and extending the Argo NetCDF format used.

Appendix A. Differences between ENACT and ENSEMBLES systems

For ENSEMBLES both WOD01 and GTSPP data were recopied from the originators websites in March/April 2005. For WOD01 no extra reports were available, but some quality corrections were included. In ENACT GTSPP was used to supplement WOD01 from 2000; it was used from 1990 in ENSEMBLES giving 5% more usable data in the early 1990s rising to 10% in the late 1990s.

In ENACT all T-4, T-6 and T-7 XBTs, and unknown XBTs less than 840 m deep, had the rescaling of depths described in section 3.1 applied. Thus some reports from 1995 onwards were 'double corrected'. Tests suggest that this caused mean XBT temperatures to be warm by about 0.17°C between 200 and 500 m.

For ENSEMBLES several changes were made to the track check – the exclusion criteria (section 3.3) were changed and the final stage (Appendix C) added. The net effect was to reject fewer observations, especially those with duplicate times. The constant value check for temperature was relaxed (section 3.2) and the prior rejects (section 3.7) were added, including the rejection of XBT data below 1000 m.

The temperature background error estimates used in ENACT were recalculated. More details of the changes are available in the EN2_v1a_changes document available via www.hadobs.org, the latest version of the observations is also available there.

Appendix B. Spike and step check

There is a preliminary rejection for tropical temperatures, above 1000 m depth, less than 1° C.
The temperature and salinity tolerances (TTol and STol) to be used at each depth are calculated (see Table B1). Then vertical differences of the temperature and salinity profiles are formed, where level *k* is at most 50

m deeper than level *k-1* (100 m where level *k* is 350 m or more deep): $DT_k = T_k - T_{k-1}$

A) If either $|DT_{k-1}| > TTol$ or $|DT_k| > TTol$ and $|DT_{k-1} + DT_k| < 0.5 \times TTol$

then T_{k-2} and T_k are in good agreement with each other, but T_{k-1} is rejected as a spike.

B) There is also a check for sharp, but smaller amplitude, temperature spikes. If

i) $|DT_{k-1}| > 0.5 \times TTol$ or $|DT_k| > 0.5 \times TTol$

ii) at least one of the vertical gradients is larger in magnitude than 0.05 °C/m and

iii) $|DT_{k-1} + DT_k| < 0.25|DT_{k-1} - DT_k|$

then T_{k-1} is rejected as a spike.

C) Remaining values of DT_{k-1} with magnitude larger than TTol represent large steps in the profile, both T_{k-2} and T_{k-1} are flagged as suspect unless one of the following conditions applies:

i) T_{k-1} agrees to within 0.5*TTol with a value interpolated from T_{k-2} and T_k

ii) for depths of 250 m or less if $0 > DT_{k-1} > -3 \times TTol$ (sharp thermocline)

iii) if it is the last DT then only the last temperature in the profile is flagged as suspect

Also if the last temperature value is zero then it is flagged as suspect (suspect values can be reprieved by the background check). If there are four or more temperature spikes or steps detected then the whole report (salinity as well) is rejected.

The salinity spike check is similar but simpler, consisting of checks analogous to A) and C), but without condition ii). If a temperature spike is detected the corresponding salinity value is automatically rejected - the temperature check is more sensitive and there is often a small "blip" in the salinity at the same level - possibly caused by the use of the temperature in deriving salinity from conductivity.

Depth (m)	TTol (°C)	STol
0	5.0	1.0
200 (*)	5.0	1.0
300 (*)	2.5	0.2
500	2.0	0.2
600	1.5	0.2

Table B1 Tolerances for the spike and step check. Within 20° of the equator the depths marked (*) are replaced by 300 and 400 m. For this upper transition the Tol values are interpolated linearly in depth. The 500 and 600 m changes in are implemented as step changes, and the 600 m values used at all deeper levels.

Appendix C. Track check algorithm

As described in section 3.3 reports from each ‘ship track’ are sorted by date and time. To allow for rounding of both position and time the implied speed between observation K-1 and K is calculated as:

Speed(K) = (Dist(K)-0.5*DistRes)/MAX(DTime,TimeRes)

where DistRes = 20000 m (20 km) and TimeRes = 600 seconds. Where both displacements are larger than DistRes then Angle(K), the change in angle at K (going from K-1 to K to K+1) is calculated, this is approximately 0° for a great circle track.

The largest speed in the whole track is found at position M. If Speed(M) is larger than MaxSpeed, or if it is larger than 0.8*MaxSpeed and either Angle(M-1) or Angle(M) is greater than 90° then one of the positions M-1 or M is deemed to be wrong. The checks below determine which to reject, the offending report(s) is (are) omitted from the sequence, the distances and angles are recalculated and the process repeated until there are no excessive speeds in the track. If more than half of the reports in the track have been rejected then the whole track is rejected.

Speed(M) >SpeedTol: deciding between M-1 and M.

If test a, b etc is positive then none of the later tests are applied (except for i), broadly speaking the tests below show diminishing returns – and very few cases reach the later tests.
a) If M=2 there is less checking that can be done. If the implied speed between 1 and 3 is less than SpeedTol and either Speed(3) > SpeedTol or Angle(3) > 45° then report 2 is rejected, otherwise report 1 is rejected. An analogous test is used if M is the last report in the track.
b) If Speed(M-1) > SpeedTol then reject M-1; if Speed(M+1) > SpeedTol then reject M.
c) If (implied speed between M-1 and M+1) > SpeedTol then reject M-1;
if (implied speed between M-2 and M) > SpeedTol then reject M.
d) If Angle(M-1) > 45 + Angle(M) then reject M-1;
if Angle(M) > 45 + Angle(M-1) then reject M.
e) If Angle(M-2) 45 and Angle(M-2) Angle(M+1) then reject M-1;
else if Angle(M+1) 45 then reject M.
f) If Speed(M-1) < MIN(Speed(M+1),0.5*MeanSpeed) then reject M-1;
if Speed(M+1) < MIN(Speed(M-1),0.5*MeanSpeed) then reject M.
MeanSpeed is calculated excluding intervals less than one hour and speeds over MaxSpeed.
g) Dist1 is calculated as the distance M-2 to M to M+1;
Dist2 is calculated as the distance M-2 to M-1 to M+1.
DistTol is the larger of DistRes and 0.1*(distance M-2 to M-1 to M to M+1).
If Dist1 < Dist2 - DistTol then reject M-1;
if Dist2 < Dist1 - DistTol then reject M.
h) PD1 = Dist(M-1)/Dist2, this is the ratio of the distance from M-2 to M-1 over the distance from M-2 to M+1 omitting M;
PD2 is analogous: the distance from M-2 to M over the distance from M-2 to M+1 omitting M-1
PT1 (PT2) is the time from M-2 to M-1 (M) over the time from M-2 to M+1; for smooth motion the ratios PD1 and PT1 (also PD2 and PT2) should be approximately equal.
If ABS(PD1-PT1) 0.1+ABS(PD2-PT2) then reject M-1;
if ABS(PD2-PT2) 0.1+ABS(PD1-PT1) then reject M.
i) If none of the above tests have been positive or if removing the rejected report, Krej, leaves an excessive speed between Krej-1 and Krej+1 then both M-1 and M are rejected. This can happen with a fairly symmetric situation where M-2 and M-1 are consistent with each other and M and M+1 are also consistent with each other but not with the first two.

There is a final stage which allows reinstatement of flagged reports if they are consistent with the unflagged reports – usually because two adjacent bad positions caused extra rejections under i). Reports with duplicate times are only checked in this final stage and allowance is made that the reported time may be wrong.

References

Bacon, S., Centurioni, L.R., and Gould, W.J., 2001: The evaluation of salinity measurements from PALACE floats. Journal of Atmospheric and Oceanic Technology 18, 1258-1266
Bell M.J., Forbes R.M., and Hines A., 2000: Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. J. Mar. Syst. 25, 1-22
Budeus, G., and Krause, G., 1993: On-cruise calibration of XBT probes. Deep-Sea Research I 40, 1359-1363
Conkright, M.E., Antonov, J.I., Baranova, O., Boyer, T.P., Garcia, H.E., Gelfeld, R., Johnson, D., Locarnini, R.A., Murphy, P.P., O'Brien, T.D., Smolyar, I., Stephens, C., 2002: World Ocean Database 2001, Volume 1: Introduction. Ed: Sydney Levitus, NOAA Atlas NESDIS 42, U.S. Government Printing Office, Washington, D.C., 167 pp.
Davey, M. ed, 2005: ENACT Final Report. Available from www.hadobs.org
Davis, R.E., Sherman, J.T., and Dufour, J., 2001: Profiling ALACEs and other advances in autonomous subsurface floats. Journal of Atmospheric and Oceanic Technology 18, 982-993

Emery, W.J., and Thomson, R.E., 2001: Data Analysis Methods in Physical Oceanography. Elsevier. xvi + 638 pp.
Hanawa, K., Rual, P., Bailey, R., Sy, A., and Szabados, M., 1995: A new depth-time equation for Sippican or TSK T-7, T-6 and T-4 expendable bathythermographs (XBT). Deep-Sea Research I 42, 1423-1451.
Hayes, S.P., Mangum, L.J., Picaut, J., Sumi, A., and Takeuchi, K., 1991: TOGA-TAO: A moored buoy array for real-time measurements in the tropical Pacific Ocean. Bull. Amer. Meteor. Soc., 72, 339-347
Heinmiller, R.H., Ebbesmayer, C.C., Taft, B.A., Olson, D.B., and Nikitin, O.P., 1983: Systematic-errors in expendable bathythermograph (XBT) profiles. . Deep-Sea Research 30, 1185-1197
Ingleby, N.B., 2001: The Statistical Structure of Forecast Errors and its Representation in The Met. Office Global 3D-Var. Quarterly Journal of Roy. Met. Soc., 127, 209-231, 2001.
Ingleby, N.B., and Lorenc, A.C., 1993: Bayesian quality control using multivariate normal distributions. Quarterly Journal of Roy. Met. Soc., 119, 1195-1225
Johnson, G.C., Sloyan, B.M., Kessler, W.S., and McTaggart, K.E., 2002: Direct measurements of upper ocean currents and water properties across the tropical Pacific during the 1990s. Progress in Oceanography 52 (1): 31-61
Kessler, W.S., Spillane, M.C., McPhaden, M.J., and Harrison, D.E., 1996: Scales of variability in the equatorial Pacific inferred from the tropical atmosphere-ocean buoy array J Climate 9 (12): 2999-3024
Kizu, S., and Hanawa, K., 2002a: Start-up transient of XBT measurement. Deep-Sea Res., 49(5), 935-940.
Kizu, S., and Hanawa, K., 2002b: Recorder-dependent temperature error of expendable bathythermograph. J. Oceanogr., 58(2), 469-476.
Kizu, S., Yoritaka, H., and Hanawa, K., 2005a: A new fall-rate equation for T-5 expendable bathythermograph (XBT) by TSK. J. Oceanogr., 61, 115-121
Kizu, S., Ito, S., and Watanabe, T., 2005b: Inter-manufacturer difference and temperature dependency of the fall-rate of T-5 expendable bathythermograph. J. Oceanogr. 61 (5): 905-912
Levitus, S., 1982: Climatological Atlas of the World Ocean, NOAA Professional Paper 13, U.S. Government Printing Office, Washington, D.C.
Lewis, E.L., and Perkin, R.G., 1981: The practical salinity scale 1978: conversion of existing data. Deep-Sea Research, 28A, 307-328
Lorenc, A.C., and Hammon, O., 1988: Objective quality control of observations using Bayesian methods. Theory, and a practical implementation. Quarterly Journal of Roy. Met. Soc., 114, 515-543
Martin, M.J., Bell, M.J. and Hines, A., 2002: Estimation of three-dimensional error covariance statistics for an ocean assimilation system. Ocean Applications Tech. Note 30. Met Office, FitzRoy Road, Exeter, EX1 3PB, UK.
McDougall, T.J., Jackett, D.R., Wright, D.G., Feistel R., 2003: Accurate and computationally efficient algorithms for potential temperature and density of seawater. Journal of Atmospheric and Oceanic Technology 20 (5): 730-741
McPhaden, M.J., Busalacchi, A.J., Cheney R., Donguy, J-R., Gage, K.S., Halpern, D., Ji, M., Julian, P., Meyers, G., Mitchum, G.T., Niiler, P.P., Picaut, J., Reynolds, R.W., Smith, N. and Takeuchi, K., 1998: The tropical ocean global atmosphere observing system: A decade of progress. J Geophys Res. 103 (C7): 14169-14240
Roemmich, D. and Cornuelle, B., 1987: Digitization and calibration of the expendable bathythermograph. Deep-Sea Research 34, 299-307
Smith, N.R., 1995: An improved system for tropical ocean subsurface temperature analyses, Journal of Atmospheric and Oceanic Technology 12 (4): 850-870
Stammer, D. 1997: Global characteristics of ocean variability estimated from regional TOPEX/POSEIDON altimeter measurements. Journal of Physical Oceanography, 27, 1743-1769
Thadathil, P., Saran, A.K., Gopalakrishna, V.V., Vethamony, P., Araligidad, N., Bailey, R., 2002: XBT fall rate in waters of extreme temperature: A case study in the Antarctic Ocean. Journal of Atmospheric and Oceanic Technology 19 (3): 391-396
Weaver, A.T., Vialard, J. and Anderson, D. L. T., 2003: Three- and four-dimensional variational assimilation with an ocean general circulation model of the tropical Pacific Ocean. Part 1: formulation, internal diagnostics and consistency checks. Mon. Wea. Rev., 131, 1360-1378.
Wang, G., Kleeman, R., Smith, N., and Tseitkin, F., 2002: The BMRC coupled general circulation model ENSO forecast system. Mon. Wea. Rev., 130, 975-991