For a Better XBT Bathy-message:
Onboard Quality Control, plus a New Data Reduction Method

Pierre RUAL

ORSTOM, Centre de Noumea, BP A5,
Noumea, New Caledonia

ABSTRACT

To improve the quality of the bathy-messages, sent to the Global Transmission System of the World Weather Watch via the Argos satellite, by ORSTOM - C.L.S. service Argos software onboard ships of opportunity, the following operations are implemented: *i)* a combination of a nonlinear median filter in order to despike the eXpandable BathyThermogramme profile prior to onboard quality control, *ii)* plus a linear Hanning filter, to filter out the small scale features before using the Broken stick data reduction method (that gives directly the required number of significant data point without iteration), *iii)* and a final regression fit. Only bathy-messages from temperature profiles that passed the onboard quality control, are sent to the satellite transmitter and the bathy-messages are computed only to the depth of the deepest "good" data point. So very few bathymessages are rejected later on, during the successive quality controls.

1. Introduction.

In order to improve the quality of the XBT (eXpandable BathyThermogramme) bathy-message sent to the GTS (Global Transmission System) of the World Weather Watch, an onboard quality control of the XBT temperature profile is necessary. The number of significant points constituting the bathy-message is limited (15 to 30), so it is important to use the best data reduction method to calculate the significant points. The bathy-message being used by meteorologists but also by atmosphere or ocean modelers, it is necessary to produce a bathy-message not only representing correctly the temperature profile but also representing with a good accuracy the integrated parameters such as the heat content. Dynamic height and geostrophic currents are deduced from the integrated density profile calculated by adding a temperature-salinity relation.

2. Onboard Quality Control.

Prior to any bathy-message calculation, a set of tests is applied to the temperature profile in order to know if it is worth doing.

All the test parameters given here are those used by ORSTOM for the TOGA (Tropical Ocean and Global Atmosphere) program in the tropical oceans, and they are to be modified according to the area of observation.

i) If the maximum depth of the profile is less then 90 meters, the profile is considered too short to be useful, and another launch is asked to the operator.

ii) If the profile is deeper than 200 meters, but its deepest temperature is higher than the 10 meters temperature minus 2°C, the profile is considered to have failed probably due to probe nose breakage, so that the probe thermistor instead of falling with its nominal speed, is floating in the mixed layer. This error is possible due to the fact that the probe depth is not measured but computed from the time elapsed since the probe contact with the sea surface. This test is a good example of a regional test that should be modified if the measuring area contain a region of deep water formation, where the water is mixed from the surface to a very great depth, as it is the case in the Antarctic Ocean.

If these two tests have been passed then the bathy-message is considered worth computing, but the question of its maximum depth is raised.

3. Bathy-message Depth.

The number of significant points (15-30) constituting a bathy-message is much reduced compared to the data points number of a profile (generally one to two per meter). So it is important to try to know the depth of the deepest "good" data point. If this depth becomes the maximum depth of the bathy-message then the later will make use of all its significant points to describe the "good" part of the profile, and the accuracy of the description will be the best obtainable with a given data reduction method.

A series of test are conducted to determine that maximum depth. Again the test parameters are those used by ORSTOM for the TOGA tropical ocean region, and should be modified if necessary.

i) In the first 10 meters, no tests are conducted as it is a region where there is a great variability both due to the ocean or to the instrument. For instance, large temperature change may be due to ice melting or cold rain or to a big difference between the storage temperature of the probe and the sea surface temperature.

ii) Starting from 10 meter, the profile is downward tested until the first of the following is met :

- the end of the profile,
- the preset maximum depth of the bathy-message (512 m with the Argos transmission,
- the first outside range temperature (-2°C, 32°C),
- the first temperature gradient above 3°C/meter, positive or negative,
- the first temperature inversion greater than 1.5°C in the first 200 meters or 0.5°C below.

iii) The temperature minimum of the profile above the data point under test, is kept in memory. The depth of the deepest data point where the temperature is less or equal to the temperature minimum plus 5/100°C (white noise value of an XBT temperature profile) is also kept in memory. This point is considered as the last "good" point of the profile and its depth is taken as the maximum depth of the bathy-message. This point may be several meters above the last data point under test.

4. Data Reduction Methods.

Once determined the fraction of the temperature profile used to compute the bathy-message, the question is, what kind of data reduction method will give the best results in the less possible processing time.

Until recently, with the strip chart recorders, this was made by hand, by skilled observers, but on ships of opportunity, where the operators are volunteers but not specialists, it is better to use an automatic method. The base for all these methods is that the profile is reasonably linear, between two successive significant points, but depending upon the premises, they are divided into two different groups:

i) A tolerance is given in the adjustment of the bathy-message to the full profile, and a certain number of significant points is deduced. If there are too many or not enough points, the tolerance may be modified and the process repeated.

ii) A number of significant points is given and the maximum deviation of the bathy-message to the profile is deduced. If that deviation is too big, then the process may be resumed adding more significant points.

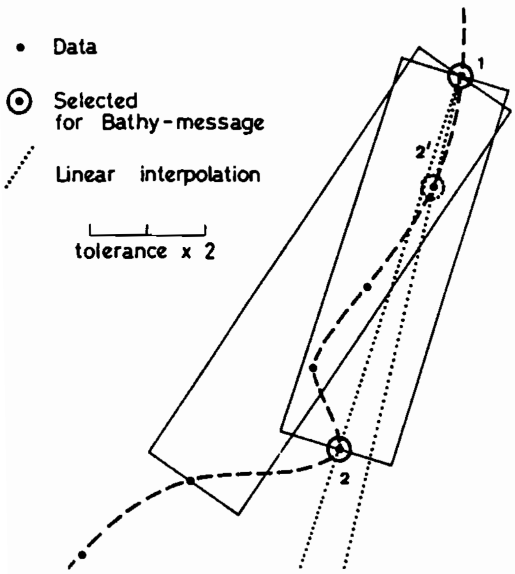

FIG.1. PIPE data reduction method.
a) Select a temperature tolerance.
b) Between 1 and 2 build a pipe with a radius equal to the tolerance.
c) Test all the data points between both ends of the pipe.
d) All of them are inside, so proceed to the data point following 2 and do again b) and c).
e) If any data point outside the pipe (here, data point 2), then select as the new significant point, either the bottom end of the previous pipe (point 2), or the data point opposite to the exit side (point 2').
f) When reaching the last data point, there is N+1 significant points. If it is too much or not enough, start again from a) with another temperature tolerance.

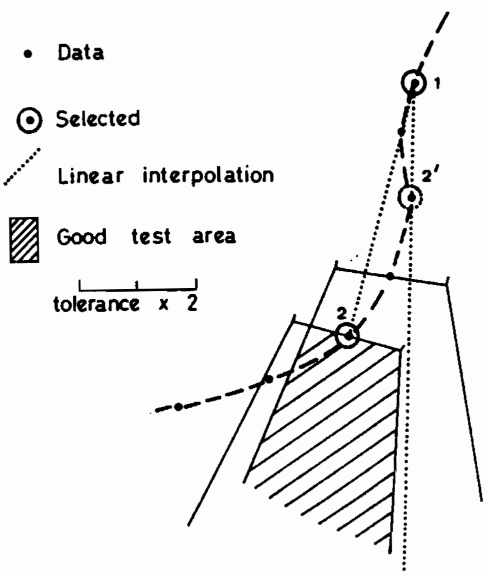

FIG.2. CONE data reduction method.
a) Select a temperature tolerance.
b) Between 1 and 2 build a cone with a base equal to twice the tolerance.
c) Test if the data point following point 2 is inside the intersection of the cone (1,2) with the previous cone.
d) If it is inside, then use it as the base of a new cone and start again b) and c).
e) If it is outside (as it is here), then select as the new significant point, either the cone base (point 2), or the data point opposite to the exit side (point 2').
f) When reaching the last data point, there is N+1 significant points. If it is too much or not enough, start again from a) with another temperature tolerance.

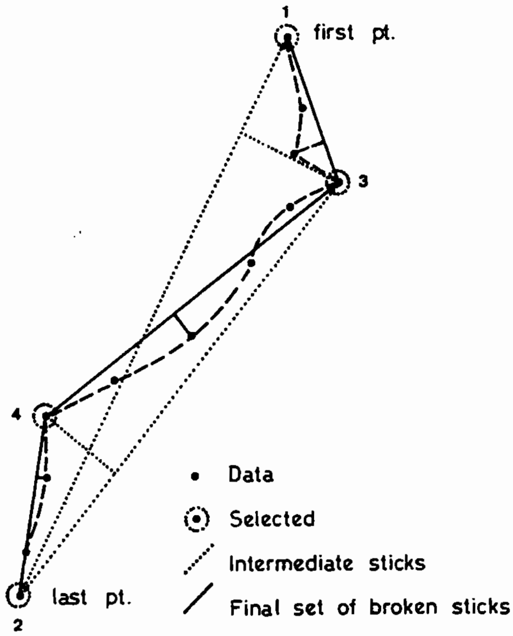

FIG.3. BROKEN STICK data reduction method.
a) Fix the number of significant points needed.
b) Draw a straight line (the stick) between the first and the last data points (1 and 2).
c) Calculate the distance between all the data points and the stick.
d) Break the stick at the data point where the distance is the greatest (point 3).
e) Having now 2 shorter sticks. Calculate the distances to the new sticks.
f) Break the stick, of all the sticks, with the data point the furthest from it (point 4).
g) Start again from b) until the selected number of significant points is reached.
Remark: Attention must be paid when breaking a stick, to check that the 2 new sticks created are not collinear with the preceding or the following stick. If so, then the intermediate point is deselected.

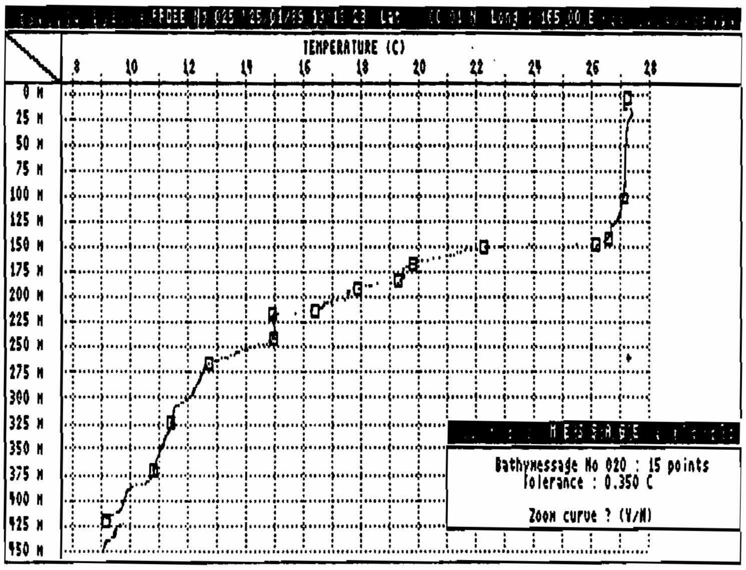

FIG.4. Bathy-message significant points. The squares represent the 15 significant points selected in 18 seconds by the Broken stick data reduction method out of a 500 data points XBT temperature profile. The maximum depth of the calculation was set to 420 meters by onboard quality control, because the temperature inversion below it, is considered too big, higher than half a degree.

a) Methods.

1) Pipe and Cone methods.

There are at least two methods in which the tolerance is given (Frachon 1987), the basic Pipe method, and the more controversial but also more rapid Cone method.

The Pipe method (Fig.1) (Siess, 1982) is equivalent to a rigid pipe with a diameter twice the tolerance, pushed around the profile, each time the profile breaks trough the pipe's wall, a significant point is found in the preceding data point or in the preceding point the most opposite to the exit. The process is then iterated starting with the new significant point. This method is very slow due to the fact that each time a new pipe is build between two points, the position of all the data points, in between, must be tested to check if they are inside or outside the pipe.

The Cone method (Fig.2) (Mesecar and Wagner, 1980; Frachon, 1987) is less satisfactory theoretically but seems to give comparable results and is very rapid as there is no backward control. The last significant point is the summit of the cone, the following point is its base, with a width twice the tolerance. If the third point is within the extrapolation of the cone, then this point is used for the base of a new cone and the next point is tested. There is one restriction, the next cone must always be inside the preceding cone, so if they intersect, only the intersection of the two cones is considered as the new cone (otherwise, it may be possible to mistake a large circle for a straight line!). When the following point is outside the cone then, as with the pipe method, a new significant point is found in the base point or in the preceding point the most opposite to the exit.

2) Broken Stick method.

The Broken Stick method (Fig.3) (Frachon, 1987; Kerr, 1984) is the only method that suppose a preset number of significant data points. Its basis is very different from the first two methods. A straight line (the stick) is drawn between the first and the last points of the profile. The distance between all the data points and the stick is computed and the stick is broken into two pieces at the data point the furthest from the stick. This point becomes a new significant point and the process is repeated, until the given number of significant points is reached. Each stick has a data point the furthest from it, but only one stick is broken at a given time, the stick containing the data point, the furthest of all the furthest data points. At the end of the process the remaining maximum distance of the furthest data point is a measure of the quality of the method and is equivalent to the tolerance of the previous methods. This method is more rapid than the Pipe method but less than the Cone method.

b) Bathy-message problem.

The GTS bathy-message format as well as most of the satellite messages, imposes a maximum length message. In order to make the best use of a message, it is best to fix the maximum number of significant points that fits into a message. The number of significant points being fixed, the obvious method to use is the Broken Stick method (Fig.4). When using one of the other methods, one has to iterate them, and to change the tolerance, until the number of significant points falls within a given range. These iterations may take a long time even with the very rapid Cone method. The empirical relation between the tolerance and the number of significant points in an XBT temperature profile, is nearly hyperbolic. Using an hyperbolic interpolation reduce a lot the number of iterations, but even then, for a 500 data points temperature profile, with a Personal Computer, the mean time to compute a 15 point bathy-message is 18 seconds with the Broken Stick method (+/- 1 second and no iteration), 45 seconds with the Cone method (9 to 90 seconds, with 9 seconds/iteration), and 2.5 minutes with the best dichotomic Pipe method (0.5 to 5 minutes, with 0.5 minute/iteration).

CRUISE : SURTROPAC 09 JAN. 88

50 Drops / 46 Bathymessage CORIOLIS

Method	Heat Content (Mean temperature)	|Error| Mean temperature	σ (profiles)	|Error Max| (profiles)
Cone	17.557°C	.043°C	.138°C	.404°C
Broken Stick		.032°C	.125°C	.362°C
BS + Regression		.004°C	.087°C	.314°C

TAB.1. Errors between the full XBT temperature profile and the bathy-messages computed using the Cone or the Broken stick method. On the third line, was added to the later a regression fit. A pseudo heat content is given in the first column, the mean of the amplitude of its error is given in the second column. The third and fourth columns present the mean of the temperature standard deviation and the mean amplitude of the maximum temperature error.

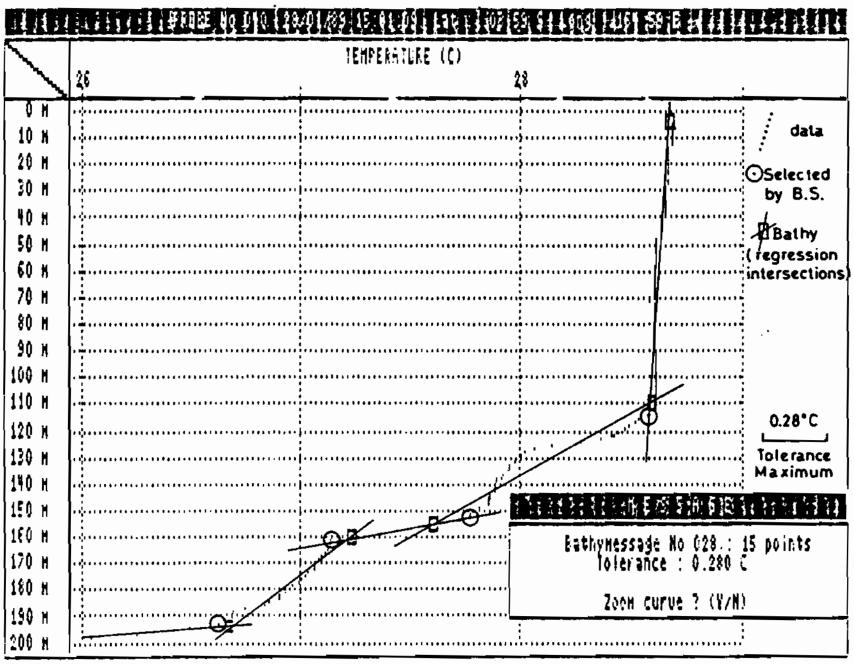

FIG.5. Linear regression fit. The open circles are the significant data points selected by a data reduction method. The squares are the intersections of the regression fit segments constructed from all the data points between two successive significant data points. The straight lines drawn between the squares are a better fit to the curve than the lines between the open circles so the regression lines intersections are chosen as new significant points.

The Broken Stick method is not only quicker but also better, as shown in Table 1, according to a study based on more than 50 temperature profiles and 46 bathy-messages. It results in a mean standard deviation and a mean maximum deviation 10% better than those of the Cone method. The mean heat content itself is 25% better. The Pipe method being too long, was not considered in this case, but its results are very close to the Cone method as seen in a previous study on a few profiles.

c) Final tests.

The final tests purpose is to verify the quality of the data reduction.

1) Pipe and Cone methods.

If the first pass with a standard tolerance does not give the right number of points, another test pass is made using the tolerance maximum (minimum) if the number of significant points is too large (too small). This will determine if it is possible to calculate the right number of points. If the resulting number of significant data points is too small (too large), then hyperbolic interpolations may converge to the required range. In the other case, the bathy-message can not be computed.

2) Broken Stick method.

After each stick breakage, the new maximum deviation between the profile and the bathy-message is computed and is checked against the tolerance minimum (generally, the white noise of the profile). If it is below then the computation stops prematurely, and if the final number of significant points is below the minimum number of points, the bathy-message is considered bad and not sent or recorded. When this is not the case, at the end of the computation, the bathy-message has the optimum number of significant data points, but its maximum deviation from the profile is checked against the tolerance maximum. If it is above then the computation is resumed till the maximum deviation is below the tolerance maximum or till the maximum number of points is reached. In this later case the bathy-message is considered bad and not sent or recorded.

Generally (Mesecar and Wagner, 1980), the tolerance maximum is 1°C and the minimum is 5/100°C, the white noise of the XBT profiles. For the GTS bathy-message format the optimum number of significant points is 20, for an Argos transmitted bathy-message it is 15. The minimum number of points is often 50% of the optimum number, and the maximum number is 10% to 20 % higher. In the case of an Argos transmission, they are respectively, 9 and 17 points, but no more than 15 points are actually transmitted because this is not possible, due to the short length of the Argos messages.

5. Improvements.

a) Linear Regression Fit .

We now have a number of significant data points belonging to the profile, and in between, a certain number of points, per definition, more or less linearly positioned. Then why not replacing these segments of curve by their linear regression fit, and use as new significant points, the intersections of the regression lines (Fig.5).

A test on more than a 150 profiles, shows in Table 2 a drastic improvement, the mean standard deviation is reduced by 30%, the mean maximum error is reduced by 13%, but more important for the modelers, the error in the integrated parameters, as the mean heat content, is reduced by an order of magnitude.

The only problem encountered while using the regression fit is when two successive regression segments are nearly parallel (Fig.6). In that case their intersection may be far from the curve and the solution is to eliminate the intersection as a significant point and to keep the significant data point.

CRUISE : ACT 9 JUL-AUG 88

105 Drops / 78 Bathymessage

Method	Heat Content (Mean temperature)	|Error| Mean temperature	σ (profiles)	|Error Max (profiles)
Cone	17.098°C	.030°C	.111°C	.339°C
Broken Stick + Regression		.002°C	.068°C	.243°C

TAB.2. Same as Table 1 for another cruise, but only with the results of the Cone method or the Broken stick plus regression fit method.

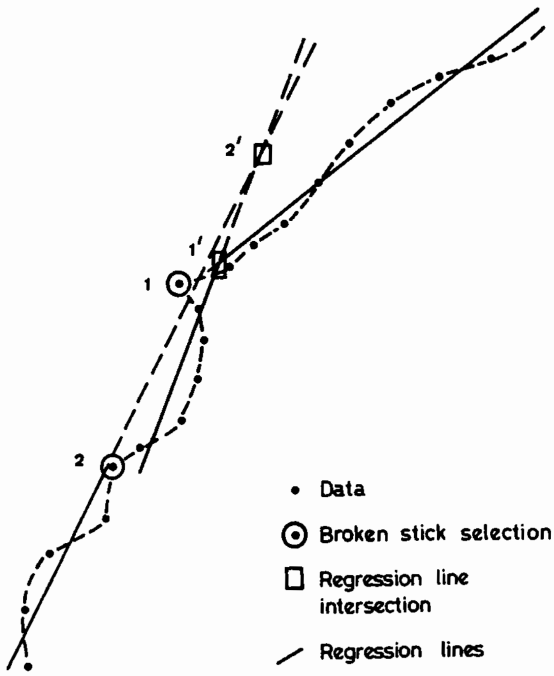

FIG.6. Quasi parallel regression lines problem. Data points 1 and 2 have been selected as significant points by some data reduction method. Points 1' and 2' are the intersections of the regression lines constructed using all the data points between two significant points. Point 1' is a better representation of the local curve and will replace the data point 1 as significant point. The two following regression segments are almost parallel. Point 2' is above 1' or 1 and is very far from the curve, so point 2' is abandoned and point 2 is kept as a significant point.

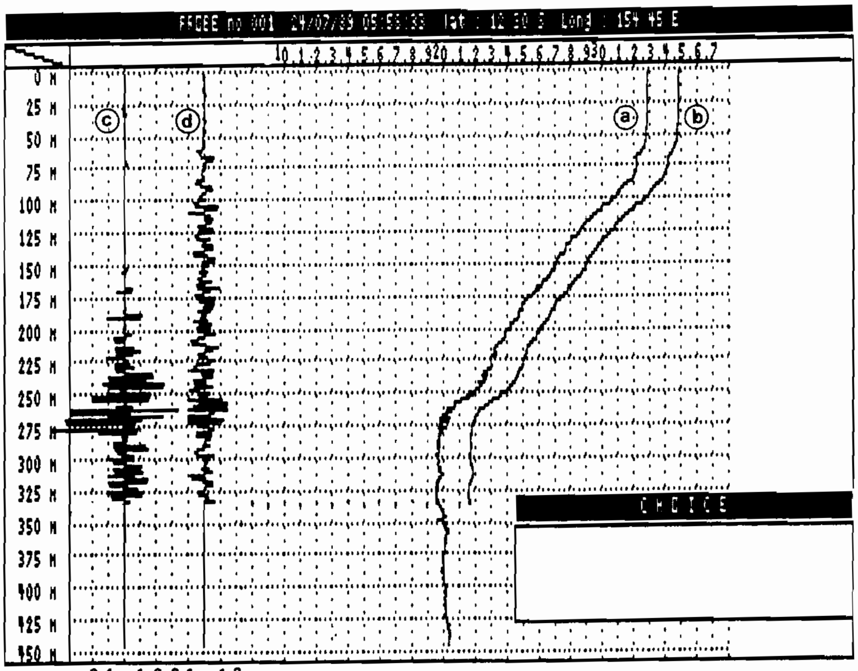

FIG.7. Median and Hanning filter effect on a temperature profile. Profile *a* is the original profile, profile *b* is the filtered profile, offset by 2°C, used to calculate the bathy-message down to 333 meters (temperature minimum). Curve *c* shows the effect of the despiking nonlinear median filter over a window of 5 points, and curve *d* the effect (offset by 5°C) of the low pass linear Hanning filter over a window of 11 points. The scales of the last two curves are multiplied by 10.

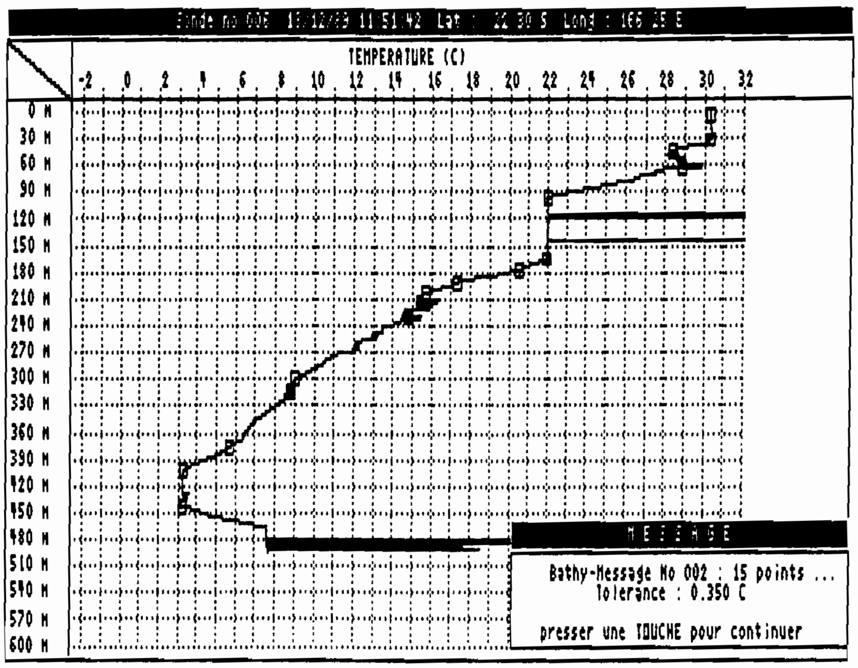

FIG.8. Improved bathy-message. Significant points are represented by the squares on that dummy temperature profile were spikes and contact problems have been amplified. Without despiking, the onboard quality control would have declared the last good data point at 118 meters and the bathy-message would have stopped at that depth. Moreover, Hanning filtering has reduced the effect of small scale features and no significant points are lost to describe them. The regression fit has also adjusted the significant points so that the vertically integrated mean temperature error is less than 2/1000°C.

b) Data Filtering.

1) Median filter.

Due to temporary defective electrical contacts, or any other cause such as radio transmission while launching, more or less isolated spikes may occur during a launch. As the search for the bathy-message maximum depth stops at the first large temperature gradient (>3°C/m), it is useful to use a despiking filter prior to the test. The nonlinear median filter (Sy, 1985) is adequate as it completely filters out the spikes with a width less than the half window, but leaves intact the data as long as the data is monotonous. A 5 meters window seems to be adequate for an XBT profile, it cuts out any spike 2 meters wide or below. A more elaborate median filter with a threshold (Brock, 1986) may also be used in order not to modify any data if it is not a spike above a given amplitude.

2) Hanning Filter.

To describe as well as possible a 500 point curve with 30 significant points or less, one wants to represent the major characteristics of the curve, leaving out the small scale features. At our disposal, now is an efficient bathy-message computing method, but it has a tendency to privilege angular points, as all data reduction methods do. So, prior to the bathy-message computation small scale features should be filtered out by a linear filter such as the Hanning filter (Fig.7) (Etienne, 1970). It is a cosine pondered running mean, a kind of Tukey filter (Matushevskiy and Prival'skiy, 1968) which has almost no secondary maximum. After comparison with CTD profiles, the best filter window seems to be 10-20 meters.

6. Conclusion.

Using a median filter (5 meters window) prior to the first quality control test, then a Hanning filter, (11 meters window) prior to the Broken stick data reduction method, and finally applying a linear regression greatly improves the bathy-message fit with the profile (Fig.8). Temperature errors are reduced by at least 40%, and errors on integrated parameters, such as the heat content, are reduced by an order of magnitude. The maximum depth of the bathy-message may also be increased by the median filter despiking. All these improvements are included in the software developed by ORSTOM and C.L.S. service Argos, for the XBT-ST Argos transmission equipment, that has been on use for more than a year on most of the French TOGA Voluntary Observing Ships network.

Furthermore, when using the Argos satellite transmission system (Table 3), the ratio between the bathy-message sent and transmitted to the GTS is very high (99%), as very few bad profiles pass the onboard quality control, and as the C.L.S. service Argos insert itself the messages, directly onto the GTS, after a rapid quality control. In the TOGA Subsurface Temperature Data Bank, the ratio between the GTS received and archived messages is also very high (99%), as most of the quality control was made during the first steps of the transmission link.

TOGA-VOS, ORSTOM PACIFIC network

January-September 1989

	Number	Step ratio	Global ratio
Probes launched	1945	-	-
Good Profiles (Noumea)	1520	78%	78%
Bathy-messages (service Argos)	1501	99%	77%
Bathy-messages (TOGA Bank)	1486	99%	76%

TAB.3. TOGA-VOS ORSTOM XBT Pacific network efficiency. The 8 ships were equipped with C.L.S.-ARGOS XBT-ST satellite transmission equipment. The second column represents the success rate between two successive lines, the third column, the global success rate between the number of probes launched and the data locally received and archived. The second line indicates the number of XBT profiles received in delayed mode in Noumea, manually quality controlled and archived. the third and fourth lines represent the real time bathy-messages received and archived after quality control, in Toulouse Argos center and in Brest TOGA subsurface temperature data bank.

Acknowledgments. This work has been possible, due to the help of the SURTROPAC ORSTOM Noumea group for the TOGA XBT Voluntary Observing Ship program, and the Toulouse based C.L.S.- service ARGOS group. The Sippican probes were funded by the US TOGA Office under an agreement with the French TOGA Representative.

REFERENCES

Brock, F.V., 1986: A Nonlinear Filter to Remove Impulse Noise from Meteorological Data. *J. Atm. Ocean. Tech.*, **3**, 51-58.

Etienne, A., 1970: Etude Spectrale des Ondes Internes et de la Turbulence, Filtrage des données. *Cahiers du COEC*, **22**, 7, 661-664.

Frachon, I.A., 1987: Réduction des Sondages bathythermiques. Service Hydrographique et Océanographique de la Marine, EPSHOM/E/OC/NP, Report N° 449, 38 pp.

Kerr, G., 1984: Chords: a New Temperature or Sound Speed Profile Thinning Algorithm. NORDA, Technical Note 272.

Mesecar, R. and J. Wagner, 1980: Instruction Manual for XBT Digital Recording and Display System, Model 791. Oregon State University Tech. Planning and Development Group, School of Oceanography. 43-112

Matushevskiy, G.V. and V.Ye. Prival'skiy, 1968: The Use of Filters in Time Series in Hydrometeorology. *Oceanology, Methods and Instruments*, **8**, 3, 401-411.

Siess, J., 1982: Algorithme des Secteurs Emboités. Service Hydrographique et Océanographique de la Marine, EPSHOM/E/OC/NP. Report

Sy, A., 1985: An Alternative Editing Technique for Oceanographic data. *Deep Sea Research*, **32**, 12, 1591-1599.

WESTERN PACIFIC INTERNATIONAL MEETING AND WORKSHOP ON TOGA COARE

Nouméa, New Caledonia
May 24-30, 1989

PROCEEDINGS

edited by

Joël Picaut *
Roger Lukas **
Thierry Delcroix *

* ORSTOM, Nouméa, New Caledonia
** JIMAR, University of Hawaii, U.S.A.

INSTITUT FRANÇAIS DE RECHERCHE SCIENTIFIQUE
POUR LE DÉVELOPPEMENT EN COOPÉRATION

ORSTOM

Centre de Nouméa

TABLE OF CONTENTS

MEETING COLLECTED PAPERS

WATER MASSES, SEA SURFACE TOPOGRAPHY, AND CIRCULATION

EL NINO/SOUTHERN OSCILLATION 1986-87

MOMENTUM, HEAT, AND MOISTURE FLUXES BETWEEN ATMOSPHERE AND OCEAN

EMPIRICAL STUDIES OF ENSO AND SHORT-TERM CLIMATE VARIABILITY

MISCELLANEOUS