BERICHTE

aus dem

INSTITUT FÜR MEERESKUNDE

an der

CHRISTIAN-ALBRECHTS-UNIVERSITÄT · KIEL

Nr. 111

1983

WARMWASSERSPHÄRE

Handling and Processing of Hydrographic Data

by

Alexander Sy

- Technical Report -

Contents

Summary

This report reviews CTD data handling and the principle of processing for the profiling instrument "Multisonde" as it is presently in use at the Institut für Meereskunde in Kiel, F.R.G., in the frame of the research programme "Warmwassersphäre". An introduction to the shipboard system of measuring and logging is given. An outline of both laboratory and in situ calibrations is presented. The main subject of this report is the processing of the data. Possible sources of errors in field measurements and their influence on data accuracy are discussed and the standard processing stages are described. A specific problem lies in data errors, which inhibit a routine processing. In order to edit these data a special filter, the median filter, is introduced. Its efficiency as well as the experiences gained through its practical application is described and discussed by means of comparison with conventional techniques. The data used for these tests were collected during the experiment NORDOSTATLANTIK '81.

Zusammenfassung

Der Bericht gibt einen Überblick über die Erfassung und die Prinzipien der Aufbereitung von CTD-Daten für das profilierende Meßgerät "Multisonde", wie sie zur Zeit im Rahmen des Sonderforschungsbereichs "Warmwassersphäre" am Institut für Meereskunde in Kiel durchgeführt wird. Neben der Vorstellung des Systems, der Messung und Registrierung sowie einer kurzen Erörterung der Labor- und in situ Kalibrierung, liegt der Schwerpunkt des Berichtes bei der Datenaufbereitung. Es werden mögliche Quellen für Meßfehler und ihre Einflüsse auf die Datenqualität diskutiert, und der standardisierte Weg der Aufbereitung wird beschrieben. Ein Problem besonderer Art stellen fehlerhafte Daten dar, die aufgrund ihrer Struktur nicht routinemäßig bearbeitet werden können. Zur Behandlung dieser Daten wird ein spezieller Filter vorgestellt, der Medianfilter. Seine Arbeitsweise und die gemachten Erfahrungen werden anhand realer Daten der Meßkampagne NORDOSTATLANTIK '81 durch Vergleich mit anderen Methoden beschrieben und diskutiert.

1. INTRODUCTION

With the establishment of the SFB research programme "Warmwassersphäre" sponsored by the German Research Society (DFG), there has been a considerable increase in the amount of hydrographic data acquired in the Institut für Meereskunde (IfM), Kiel, F.R.G.. This required new plans about their handling and processing. Standardization of acquisition, calibration and processing was necessary to compare the data sets of different cruises using different equipment.

The subject of this report is to give investigators in our project an overview about our handling of the hydrographic data. It is hoped that this report will help to eliminate sources of errors. Because the measuring and logging of CTD data is largely determined by hardware devices and is thus more a matter of technical control, the main focus of this report is on the elimination of errors by means of software.

The second purpose of this report is to give subsequent users of our data a presentation of our principles of acquisition and processing to make qualitative evaluation easy. For this reason this document is joined to the processed data sets which we gathered during NORDOSTATLANTIK '81, '82 (NOA) and following. The data are reported in Fahrbach et al. (1983) together with quantitative estimates of their accuracy.

2. INSTRUMENTATION AND DATA ACQUISITION

The shipboard system of measuring and logging uses parallel instrumentation, with the aim that observations continue despite system failures (fig. 1). The normal data flow goes from the CTD probe via deck's unit to the ship's computer (Data General NOVA) storing this data in physical units with a sampling rate of 16 s^{-1}. In the case that the ship's computer fails or if there is none available the raw data (frequencies) can be recorded by an analogue tape-recorder. Later on, inputting the tape-recorder like a CTD probe, this signal can be handled as usual. An improved procedure is planned whereby binary counts are directly recorded on a 9-track magnetic tape instead of recording frequencies on commercial "hifi" tape; this avoids repeating the stations time-wise during the transfer to a computer. Another back-up is possible by punching the calibrated data on paper tape with a reduced sampling rate of 0.2 s^{-1}. After the transfer of data to an institute's computer with appropriate format changes, the processing follows through five stages.

2.1 The CTD device

The in situ measurement of the hydrographic parameters pressure P, temperature T and conductivity C is carried out using the "Kiel Multisonde" which was developed by the "Institut für Angewandte Physik" in Kiel, F.R.G.. The "Multisonde" is commercially produced by "Meerestechnik-Elektronik GmbH (ME)" in Trappenkamp, F.R.G.. For detailed information about the principle of measuring and data transmission as well as further citations see Kroebel et al. (1976) and Biermann et al. (1976).

The profiling CTD device consists of an underwater unit containing the probes which is connected with the deck's

unit by a single-conductor-armoured cable. The deck's unit provides the power supply of the probes, the conversion of frequencies into counts and physical units, and displays the measured parameters including the derived salinity. The output can be transmitted for different applications in physical units with a low data rate to a printer, analogue plotter, tape punch, and with a high data rate to a digital computer which stores counts and physical units (raw data) on 9-track magnetic tapes.

The scanning time interval of the probe is $\Delta s = 2$ ms, but the telemetry system and the number of sensors determines the minimum sampling interval Δt. It is designed for a maximum transmission rate of 128 samples per second with one probe in use or 32 samples per second when four probes are used. In the CTD devices of the IfM the sampling time is tuned to $\Delta t = 0.064$ s. The intermediate scans are averaged. During the NOA cruises we chose a sampling interval of $\Delta t = 0.1$ s when data were recorded on computer magnetic tapes. Due to the short scan time and the averaging, a quasi-simultaneous measurement of temperature, conductivity and pressure can be assumed.

The technical data of the probes are summarized in table 1. It has to be pointed out that the accuracy of temperature and conductivity as specified in table 1 depends in practice on the quality of the laboratory calibration of the manufacturer. Besides the sensor response characteristics the calibration is the limiting factor on the accuracy.

Table 1: Technical data of the "Multisonde" according to manufacturer's declaration in the user's manual.

Pressure: Principle   Strain-Gauge Pressure Cell
              Range   0 - 6000 dbar
              Resolution   16 bit $\cong$ 0.2 dbar
              Accuracy   0.35 % of range

Temperature: Principle   Platinum Resistance
              Range   -2 °C - +35 °C
              Time lag   60 ms (without protecting
                                     sheath)
              Resolution   16 bit $\cong$ 1 mK
              Long Term Stability ±5 mK/0.5 y
              (Accuracy   ±5 mK)

Conductivity: Principle   Symmetric Electrode Cell
              Range   5 - 55 mS/cm
              Resolution   16 bit $\cong$ 1 µS/cm
              Long Term Stability ±10 µS/cm/0.5 y
              (Accuracy   ±5 µS/cm)

2.2 Calibration

The conversion from frequencies to physical units is carried out by an externally programmable microcomputer in the deck's unit or by the special operational system "DT7" during the recording on magnetic tapes with the ship's computer. "DT7" was developed by Rathlev (1981).

The required coefficients are obtained by laboratory calibration. By means of a least square fit the coefficients a_{ij} of the polynomial

$$y_j = a_{0j} + a_{1j}x_j + a_{2j}x_j^2 + \ldots$$

are determined, where the index j represents pressure, temperature or conductivity

and $\quad y_j = (P, T, \frac{1}{C})$

$$x = \tan(\frac{\pi}{2}(1 - \frac{N}{N_0}))$$

for the "tangent" CTD devices (MS01, MS02, MS35)

or $\quad y_j = (P, T, C)$

$\qquad x = N - N_0$

for the new "linear" Multisonde (MS45)

with $\quad N_0 = 2^{15} = 32768$

$\qquad N$ = actual count.

This polynomial approximates an ensemble of about 15 data pairs consisting of counts from the deck's unit and the corresponding readings of a substandard. The units are dbar for pressure, °C for temperature, mS/cm for conductivity and 10^{-3} for the salinity derived from the three previous parameters according to the practical salinity definition of 1978 (Unesco, 1981).

The calibration of pressure is realized by means of a piston gauge. The rms error of the approximation, with typical magnitude of 1 - 3 dbar, is smaller than the error of 0.35 % of the total range quoted by the manufacturer and caused by thermal expansion and hysteresis.

The temperature is calibrated using various methods. A 400 l salt water bath containing the entire probe including the electronics together with a calibrated mercury thermometer is cooled from 34 °C down to -2 °C. The accuracy of the calibration is determined by the mercury thermometer and temperature gradients within the water bath which can be

neglected here. Hence an error of 10 mK must be accepted. This method was applied at the IfM up to 1981. A new water bath and a more appropriate technique was introduced in 1982. It uses the platinum thermometer PT100 which is compared with a mercury thermometer and triple point cells of water (0.0100 °C ± 0.5 mK) and gallium (29.772 °C ± 1 mK) (Langhof & Zenk, 1982). The accuracy of the approximation is better than the manufacturer's given value (Tab. 1).

As with temperature, the technique of conductivity calibration is still in the stage of development. The calibration is obtained for a reduced range of approximately 30 - 60 mS/cm by using a small container (2 l) with standard sea water which is inserted in a 200 l water bath. This water bath is cooled from 30 °C down to -2 °C. With a careful execution of the measurements, the main error is due to the accuracy of the mercury thermometer giving the temperature of the water bath. Hence, an error of 10 µS/cm must be expected.

The handling of conductivity calibration was improved using temperature and conductivity sensors as substandards. As a further aim using the standardization relative to triple points of water and gallium the error of the conductivity calibration should be reduced significantly.

The calculation of conductivity C = C(0,T,S) is done by using the salinity algorithm according to Fofonoff & Millard (1982) with C(0,15,35) = 42.902 mS/cm. Because there is no unique value for the conductivity of standard seawater recommended by UNESCO the above value was chosen for reasons of compatibility only.

2.3 In situ comparison

As noted by Baker (1981) a CTD system does not replace the use of water bottles. It is necessary to check the laboratory calibrations, especially the calibration drift for those calibrations made some time ago. Therefore, we carried out reference measurements for a possible correction of the data during the processing. Due to the lack of a rosette water sampler, most of the water samples were collected by separate hydrographic casts, in addition to CTD profiles. Thus, only samples from regions of small gradients, e.g. mixed layer and the depth below 2000 m could be evaluated. Often this method does not yield sufficiently good measurements and the available data are not distributed over the whole range. Hence, this procedure can only be used to correct an offset. In each error calculation this should be noted as a zero order correction only. The water samples have been analysed with a "Guildline Autosal Laboratory Salinometer Model 8400" and the corrections were applied after the recalculation of in situ conductivity. Only a rosette sampler combined with the profiling CTD device can yield satisfying in situ measurements of both temperature and salinity for monitoring the calibrations of these parameters.

The in situ pressure comparison is restricted to a zero pressure level check. For that purpose the length of the cable and the pressure indication by the CTD deck's unit are compared with a CTD devise at a depth close to the surface. A pressure check by means of a computed thermometric pressure using unprotected thermometers is possible in principle, but the error is of the same order of magnitude like that of the pressure sensor.

3. DATA PROCESSING

Data processing has to be done for different purposes. Data that are collected with different shipboard systems for measuring and logging have to be related to the same standard. Quality has to be improved by correction and removal of erroneous values. The data set has to be reduced to a manageable size without losing information. The resulting data should be instrument independent. The schematic procedure is shown in fig. 1. The use of the flow chart will vary according to the characteristics of the raw data and the expected requirements. There is no uniform procedure for all possible cases. Therefore, a problem orientated processing is recommended. That makes data processing sometimes a time consuming and laborious task.

For the large and mesoscale analysis of hydrography in the area north of the Acores along the Midatlantic Ridge, a data set with high vertical resolution is not necessary. Further, our available probes are not designed for micro-structure measurements. Therefore it was sufficient to collect data with a maximum sampling interval of $\Delta t = 0.1$ s using an instrument lowering rate of 1 - 1.5 m/s. That seems to be enough data for processing and compacting for final pressure increments of 1 dbar. The chosen pressure interval is a far higher resolution than is needed for the meso-scale analysis, and is the absolutely lowest boundary which keeps open the option of better resolution for subsequent investigations.

3.1 Sources of errors

It must always be kept in mind that table 1 contains specifications of the sensors which are determined exclusively by laboratory tests. Under field conditions various factors can restrict the accuracy. One has to

identify these additional sources of errors to improve the data quality. Some possible sources of errors are summarized in the following.

a) The differences in time constants of temperature τ_T, conductivity τ_C and pressure sensors τ_p cause a response mismatch for these sensors. The time constant τ is usually defined as that time required for the sensor to register 63 % of the amplitude of a step-like change in water property assuming an exponential rise of the response. The response of the temperature sensor is in general slow compared to the pressure sensor. The response of the conductivity cell is not instantaneous but determined by the flushing time of the cell which can be estimated as $\tau_C = 0.55 \cdot \frac{L}{v}$, with L the length of the cell and v the lowering speed (Perkin & Lewis, 1982). τ_C is of the same order of magnitude as τ_T but in general they are not equal. This causes an offset in the derived salinity in regions of constant vertical temperature gradient and overshoots (spikes) at step-like structures of temperature and conductivity (Dantzler, 1974; Scarlet, 1975; Horne & Toole, 1980). For downcast data, $\tau_T > \tau_C$ and a profile with decreasing temperature this effect leads to higher values in temperature and lower values in the calculated salinity.

A similar error occurs in salinity data due to the mismatch of sampling time, i.e. sequential instead of simultaneous measurements (Roden & Irish, 1975), or due to the spatial separation of the sensors. The last two sources of errors can be neglected in the case of the "Multisonde" but the response discrepancy due to different time constants must be compensated.

Following Horne & Toole (1980), the observed temperature is smoothed and distorted in comparison to the actual

temperature due to the finite size and the thermal mass of the sensor head. It is assumed that this effect is not relevant for our non-micro-structure measurements.

b) The specific shape of the conductivity cell leads to several error sources causing salinity spikes.

1. The heat transfer between the conductivity cell and the water within it causes an error in conductivity which depends on the temperature gradient and on the flushing time as a function of lowering rate (Scarlet, 1975).

2. A laminar instead of turbulent flow in the bore causes a trapping due to a viscous boundary layer. An incomplete water exchange in the cell can result which can produce a difference between the conductivity in the cell and that of the environment (Gregg et al., 1982).

3. The same effect due to incomplete water exchange can be caused by slow lowering rates, or due to the ship's roll or by transient blocking of the cell opening.

4. Depending on the ship's rolling and pitching, the profile does not consist of downcast data with a constant lowering rate alone, but rather there are violent changes in the speed, including velocity reversal (e.g. fig. 2a, b, c). Because the sensors are situated close to the bottom of the CTD device for undisturbed flow while descending, the upcast data are measured in the wake of the sensor housings. They should be treated like erroneous data and be discarded. To avoid these effects, especially the velocity reversal when the sea runs high, a minimum

acceptable lowering speed should be guaranteed during the measurement.

The fastest possible lowering speed for the "Multisonde" is 2 m/s. If the lowering is too fast, the form drag of the instrument body may cause lateral motion. As Gregg et al. (1982) pointed out, a distorted flow in and around the conductivity cell can be caused by a significant angle of attack introduced by instabilities of the descent, ship's drift and ambient shear.

5. According to Fofonoff et al. (1974) a coating on the inside of the conductivity cell will reduce the measured conductivity due to a change of the cell factor. Further, temperature and pressure changes can also change the cell factor. However, laboratory tests (T. Müller, pers. comm.) do not show this to be significant for the "Multisonde". A high accuracy of the in situ comparisons allows the investigator to compensate for these errors should they prove to be significant.

c) The significant knowledge of the hysteresis characteristics and the characteristics of the heat transfer to the pressure sensor causes a direct error in the pressure date which are given with the sensor specification. this uncertainty in the pressure gives rise to an uncertainty in the salinity calculation which in turn complicates the in situ comparison. This happens especially when collecting water samples by means of a rosette sampler during the ascent of the CTD (H. Peters, pers. comm.). For that reason an investigation of the hysteresis is being conducted but is not yet finished.

d) Noise in all parameters is produced by digitizing, by the electronics and by component malfunctions which can raise the noise level. Noise leads to unrealistic density inversions (Fofonoff et al., 1974; Irish & Levine, 1978).

These errors gain importance in regions of strong gradients, i.e. especially in the seasonal thermocline and of course for micro-structure measurements. Hence a good bit of attention during the processing is paid on discrimination against high or low quality data and to discard erroneous data values.

3.2 Routine processing of data of good technical quality

After data transfer from the ship's computer (NOVA) to the institute's computer (PDP-11) and translation into the institute's standard format (MK4), a data cycle consists of time t, latitude ϕ, longitude λ, depth P, temperature T, conductivity C, and salinity S. The measured parameters were archived in physical units as well as in non-calibrated counts.

Stage 1:

With a simple coarse filter using maximum and minimum bounds according to user's specification, only physically reason-able data values are accepted for further processing. The onsets of profiles are determined, i.e. all data measured during pre-cast equalization are discarded. Also, at this time, the header information is checked, corrected and completed.

If necessary, the measured data are edited by removing erroneous values in P, T and C. For the different editing techniques see 3.3 and fig. 4, 5 and 6. The use of the median filter technique MEDFIL is not recommended in stage 1

before the time lag correction is carried out. Additionally, errors in temperature should be removed before the time lag correction is applied to avoid the influence of erroneous values on the calculation of the corrected value. This is especially important for profiles with narrowly spaced groups of spikes. For the same reason the bad data points should be substituted with expected values using other methods like STADEV or MEDDIF or an interpolation technique.

Stage 2:

Assuming the validity of Newton's law of cooling, the time constant correction is done by an algorithm using first order backward differences as described by Peters (1976) whereby the true temperature T_O is approximated by

$$T_O(i) = T(i) + \frac{\Delta\tau}{\Delta t} \left(T(i) - T(i-1)\right)$$

with $T(i)$ the temperature reading at sample i and $\Delta\tau = \tau_T - \tau_C$ the effective time constant. This simple procedure should be improved for several reasons. An essential inaccuracy in time lag corrections is caused by uncertainty of time constants themselves. As the investigation of Hurst (1975) shows, the time constant of a sensor depends on several conditions, and Gregg et al. (1982) show for the Neil Brown CTD that the amplitude response of the conductivity cell is more strongly attenuated than that of the thermistor at low sensor speeds, and the reverse is true at high sensor speeds. Hence there is evidence to doubt the time constant values specified by the manufacturer. Further, the effective time constant is a function of the lowering speed.

To obtain the effective time constant value a method according to Fofonoff et al. (1974) can be used, which minimizes typical time lag induced salinity spikes. A more suitable method, the calculation of cross spectra of temperature and salinity is described by Joyce (1976).

According to the rolling and pitching of the ship the lowering speed varies and thus also the time constant of the conductivity cell. The accurate determination of the effective time constant requires a lowering rate dependent technique. For our present data a sufficient calculation of the lowering speed from pressure readings is not possible due to the resolution of the pressure sensor and therefore the time lag correction is applied for an average response only. Anyway, for data obtained in bad weather, which is not unusual in the North Atlantic, the correction is doubtful.

Additionally, the determination of the time constant by an indirect method needs a more efficient correction filter, as described by Grose (1981) or Horne & Toole (1980). Besides recursive filter techniques, a time constant correction by applying a time lag to pressure and conductivity instead of a correction of temperature is recommended to avoid noise amplification at high frequencies (Grose, 1981).

The time-lag induced bias in salinity in a region of constant temperature gradient is of no importance for our present data. In the NACW water mass, the region with the strongest mean gradient, we find $\frac{\Delta T}{\Delta z} \approx 0.01$ K/m. Assuming a temperature time constant two times larger than the 60 ms specified by the manufacturer, i.e. $\tau_T = 120$ ms and a lowering speed of 1 m/s, which corresponds to a conductivity flushing time $\tau_C = 55$ ms, one obtains an effective time lag $\Delta\tau = 65$ ms or 0.65 cycles. This results in a temperature offset of about $\delta T = +0.00065$ °C and a corresponding salinity offset $\delta S \approx -0.0005 \cdot 10^{-3}$ is found.

Next, follows the linear correction of the profiles corresponding to the in situ comparisons of P, T and C. This procedure is to be done before the recalculation of salinity and density. These two parameters provide the possibility of controlling the response of the previous adjustments,

because both are highly sensitive to errors in measured temperature and conductivity.

A first data reduction can be realized after the time lag correction by enforcing a strict monotonic sequence in pressure. This procedure reduces the amount of data up to 50 % and is recommended in an early stage for economic reasons. In addition, the above mentioned effects caused by the ship's roll are largely removed (e.g. see fig. 2d). In any case, a monotonic profile should be produced before a smoothing median filter is used (see 3.3.2).

For optimal reduction of the various influences of ship's roll, a minimum acceptable lowering rate algorithm should be introduced. Given a sampling rate of 10 s^{-1}, the slowest lowering speed compatible with the 0.2 dbar pressure resolution is 2 m/s. Our normal mean lowering speed of 1 m/s implies an oversampling with two observations occuring during one pressure interval on the average. Therefore the introduction of a minimum acceptable lowering rate is useless for our present data.

Stage 3:

With high quality data and a sufficient time lag compensation, no more spikes should be found in the salinity. Often this is not the case. That requires despiking of salinity with an appropriate spike rejection criterion as described in 3.3. The main region where salinity spikes occur is the seasonal thermocline, a domain of strong gradients. The sensitivity of this region to erroneous measurements is pointed out in 3.1. The more sophisticated technique, as recommended by Grose (1981), is obviously necessary to improve our time lag compensation.

Stage 4:

For some particular cases editing by hand is necessary. Routine editing algorithms fail for profiles with extended measuring intervals containing more than 50 % erroneous data. When such regions are not obvious, i.e. in the case of a slow drift of small intensity, discrimination between correct and wrong data can be difficult.

Stage 5:

Finally the edited data are smoothed with a moving average filter if a smoothing median filter was not used during the previous stages. Then the data set is compacted by arithmetic averaging, by interpolation or by both techniques combined. This compaction must be done with regard to aliasing, the resolution of the reduced data and the enhanced statistical dependence of adjacent data points caused by processing. Further, the subsequent user's needs have to be taken into account. If the median filter was used for spike editing, the compaction has to be done by averaging over a constant pressure interval, choosing the averaging interval with regard to the window size of the median filter. Otherwise, an interpolation to equidistant pressure intervals is necessary corresponding to the filter characteristics of previous smoothing procedures.

The final data set is archived on PDP-11 9-track magnetic tapes with MK4 formatted files with pressure P in dbar, temperature T in °C, salinity S in 10^{-3} (psu) and density as $\sigma_t = \left(\rho(S,T,P=0) - 10^3\right)$ in kg m^{-3}. Each step of the processing is documented with one comment line in addition to the header information. The header contains information such as the details of data storage, instrumentation and nautical information.

3.3 Processing of data of bad technical quality

3.3.1 General problems

Routine data processing becomes difficult in parts of the
profiles where serious disturbances contaminate the signal.
These disturbances are mainly caused by malfunction of the
instrumentation like slip ring noise, power supply
instabilities, influences of the wireless station etc. They
are difficult and sometimes impossible to remove. Even when
an adequate technique seems to remove the disturbances, a
falsification of the original signal can occur.
The measured signal can be disturbed in two ways: (1) A more
or less obvious drift alters the signal totally or in
extended parts; (2) a limited number of individual data
points are erroneous and can be identified as overshoots or
"spikes". The drift of an entire profile, which is often not
visible, can be detected and corrected by means of in situ
comparisons with rosette sampler data. The example in
fig. 8d shows a drift within a 100 dbar pressure interval
between 300 and 400 dbar in profile St. 631. Nevertheless
its trend cannot be determined and hence this drift cannot
be corrected. The whole erroneous interval must be
eliminated by hand according to stage 4, with loss of all of
the information.

The removal of spikes raises problems due to their density,
extent, size and position in the profile. Especially
narrowly spaced groups of spikes and/or spikes within areas
of strong gradients often cause considerable difficulties.
Additionally, it is desired to minimize the information
loss.

Identifying and clipping spikes on a maximum acceptable
first difference basis (e.g. MAXDIF, fig. 4) is the most
common way of de-spiking. Our experiences with processing of

data sets from the equatorial Atlantic show that despiking on a maximum difference basis is not effective in strong gradient zones like the salinity maximum. The data in fig. 3, chosen as an example to demonstrate this type of problem, were collected in 1979 with a "Bathysonde" and the spikes were caused by slip ring noise. Many correct data points can be lost by this method. Replacement of the eliminated data cycles by interpolation runs the risk of biasing the profiles, especially in gradient zones and has to be done with the greatest care. Another algorithm is based on the standard deviation (e.g. STADEV, fig. 5). It will be shown in an example that this method can produce serious errors.

3.3.2 An alternative editing technique

To overcome the disadvantages of the previously described methods and the relatively great loss of information, and to ensure a safe editing, i.e. to combine a maximum filtering of wrong data with a minimum modification of correct data, a technique based upon a median inquiry was introduced (MEDFIL and MEDDIF, see fig. 6).

A value x, which satifies $P(X < x) > \frac{1}{2}$ and $P(X > x) > \frac{1}{2}$ is called median $\tilde{x}$, with $P(X)$ the probability for the variable X. Or in other words: for a sequence of n data points $x_1, \ldots, x_n$ with n odd, the median $\tilde{x}$ is defined by the central data point after reordering the set according to the size of the data points. With a data window of width $Q = 2K + 1$, the median is $\tilde{x} = $ median $[x(J - K), \ldots, x(J), \ldots, x(J + K)]$ and $\{J, K \in \mathbb{Z}\}$. The median inquiry works like a low pass filter which suppresses noise as well as simple spikes as suggested previously by Tukey (1971). It is a well known tool in all social science statistics, used successfully in the '70s in the processing of digital

images, and is suited as well to the economical processing of one-dimensional time series. The median filter is a non-linear operation which complicates the mathematical analysis of its properties. In general only simple cases are investigated (Justusson, 1981). An example of a one-dimensional series treated with this filter is shown in fig. 7.

In contrast to conventional methods, data editing with the median filter has the following advantages:

1. The number of cycles is not reduced.

2. Gradients and edges are preserved. Optimal results are obtained in regions of strong gradients.

3. The profiles are smoothed (with MEDFIL but not with MEDDIF) with only a little blurring of exposed structures within the window.

4. The simple way of handling makes this method suitable for "on-line" processing.

The disadvantages are:

1. A general response function cannot be specified because it depends on the actual input signal.

2. Data must be compacted by arithmetic averaging because of the modified statistical dependence and because of the changed relation of the parameters within a cycle.

3. There can be a bias due to the occurrence of a dense pack of spikes, which are themselves biased in direction.

For further details of the median filter see Justusson (1981) and Tyan (1981).

In practice this method proved to be powerful not only for data processing in the laboratory at Kiel but also during operations at sea. In the acquisition system of a towed CTD device which is in use at the IfM, Kiel, the median filter is applied "on-line" with $Q = 5$, followed by averaging over 5 data points (Horch, pers. comm.).

It was observed that within an individual CTD profile between 40 % and 70 % of the data points are re-arranged by the median filter. This depends on the window size and the noise in the data. For suppressing spikes, the window size has to be extended at least over $2N + 1$ data points, with N the width of the spike. The number of wrong data points must not exceed N within the window. That means, there must be more correct than erroneous data points within an interval of window size. To avoid the disadvantages mentioned in number 2, of the above listing, the median filter was combined with a maximum difference inquiry (MEDDIF, fig. 6). In this way smoothing is avoided without loosing the despiking advantages of the filter.

3.3.3 Comparison of editing techniques

To demonstrate the efficiency of the editing techniques, three typical profiles with different types of data errors were chosen from a cruise in 1981. As a starting point for the discussion, the sequence of temperature, conductivity, salinity and density is shown in figs. 8a-d. Processing has been carried out to stage 1 of fig. 1 with editing of pressure only, and additionally, a linear correction of P, T, C was applied. The first profile (St. 614 in fig. 8) is an example of a record with essentially no errors. The

second (St. 631) shows numerous spikes in temperature together with a drift in conductivity in the depth interval between 300 m and 400 m. This drift in conductivity becomes visible only in salinity and density. In the third profile (St. 661) we find scattered groups of spikes of variable extent.

The second profile was chosen as a worst-case example because the spikes are not located in a gradient zone. Thus, results which are not obtained by the method based upon maximum acceptable first differences (MAXDIF) can be compared with results which are obtained by MAXDIF. This method works best when no strong gradients are present, and is used as a standard. In contrast to MAXDIF, the median filter method MEDFIL shows best results in exactly these strong gradient regimes, because a monotonic sequence is invariant to a median filter. Therefore a possible inadequacy of MEDFIL can be detected through a comparison with MAXDIF.

Fig. 9a shows an enlarged cut from this spiky interval of the temperature record of St. 631. A further enlargement (fig. 9b) reveals individual data points. Both temperature and conductivity are edited with four techniques: MEDFIL, MEDDIF, STADEV and MAXDIF. The results of the first run are shown in fig. 10 and fig. 11a, b, c, d. The parameters used are listed in table 2.

The first run produces some apparent differences using the different methods. The strong smoothing effect of MEDFIL is evident, but according to Justusson (1981) the suppressing of white noise should be smaller than with a moving average filter with the same number of filter weights. In contrast to MAXDIF both median filters did not remove the spikes completely. This property of MEDFIL is typical when there are narrowly spaced spikes. An enlargement of window size

by, say, 2 or 4 data points would not improve this result significantly but will give more signal distortion. A second run even with a smaller window size will do it (see tab. 2). That means the median filter is now sufficiently stationary. This iteration can be done until a sequence is obtained which is invariant to further median filtering. This will be reached after a few repetitions with only small rearrangements after the first run. Obviously the stationarity condition depends on the chosen window size as well as upon the data itself. In most cases two runs seem to be sufficient.

Table 2: Parameters for editing runs of St. 631

	1. run		2. run	
A: MEDFIL	T, C:	Q = 7	T, C:	Q = 5
B: MEDDIF	T, C:	Q = 7 ΔT = .05 ΔC = .05	T, C:	Q = 5 ΔT = .05 ΔC = .05
C: STADEV	T, C:	Q = 49 C = 3	T, C:	Q = 25 C = 3
D: MAXDIF	T, C:	Q = 7 ΔT = .1 ΔC = .1	T, C:	Q = 5 ΔT = .05 ΔC = .05

After the second run, the time lag correction was carried out, a strict monotony in pressure enforced, the salinity recalculated, the offset corrected and the density recalculated. If possible, i.e. if permitted by the data quality, the time lag correction should be done before

median filtering. If data quality is poor and editing must precede time lag correction, then the editing can be done by MEDDIF, by discarding instead of replacing erroneous data points. Due to the (expected) small offset error, the time lag correction after median filtering was admissable. Because gradients are not modified, the sequence of the data is not altered and therefore, the time lag correction can be applied. The resulting series are shown in fig. 12a-d for temperature, fig. 13a-d for conductivity, fig. 14a-d for salinity and fig. 15a-d for density. An enlarged cut of the processed temperature from the spiky part is shown in figs. 16 and 17 to compare all applied methods. Fig. 17 shows the individual data points.

It is evident that STADEV produces the greatest deviation compared to the other profiles in this spiky region. The occurrence of asymmetrical, narrowly spaced spikes biases the mean profile when using STADEV. That makes this method unsuitable for the chosen case. A reduction of this error should be possible, in fact, by a stronger smoothing and wider boundaries for the admitted standard deviation, but this would require a too extensive procedure for processing of numerous profiles. Further deformation of the salinity in the thermocline (fig. 14c) is to be seen and accordingly in the density (fig. 15c). This is caused by blurring of the step-like temperature using the smoothing technique. With this example the sensitivity on the choice of the appropriate parameters is shown. Therefore STADEV cannot be recommended for routine processing of a large amount of data.

The above mentioned biasing effect of narrowly spaced asymmetric spikes is true for MEDFIL as well, but the mean difference compared to MAXDIF is smaller than 0.02 K (fig. 17). In this figure the profile obtained with MEDDIF is not plotted because this curve coincides with that from

MAXDIF. There is only one difference between them. With MEDDIF the number of cycles is maintained while using MAXDIF about 1/3 of the cycles were lost in the interval between 350 and 360 dbar. It should be noted also that the use of MEDDIF without replacement of erroneous data points by the median of an interval causes the same result as MAXDIF.

Salinity was edited in the same way as temperature and conductivity in stage 2 (see tab. 2). Consequently density was recalculated. The results are presented in fig. 18b. The salinity calculated from the processed pressure, temperature and conductivity before editing is shown in fig. 18a for comparison.

Even though narrowly spaced groups of spikes can be satisfactorily removed, the density profiles show (fig. 15a-d) that the remaining relative salinity maximum of station 631 is not real. It can be related to the above mentioned drift in conductivity. Non-correctable errors of this kind are removable by hand editing only.

In figs. 19a-d and 20a-d, the final data of salinity and density are shown. All profiles are averaged within 2 dbar pressure intervals. The profiles processed with STADEV and MAXDIF are additionally smoothed with a five-point moving average filter to get better comparability with the results from MEDFIL.

In the seasonal thermocline the median filter technique is superior to the maximum acceptable difference method. This can be expected from theoretical considerations. However, it must be noted that this conclusion may not be generally true. The spectra of the upper 200 m show no significant differences between MEDFIL and MAXDIF. Discrepancies obtained by using different techniques are visible in the TS-relations of station 614 plotted in figs. 21 and 22a-d.

The median filter technique seems to produce the best result.

For the comparison of raw data and final data of the profile at station 614, spectra are shown in fig. 23a,b for both temperature and salinity. The raw data are interpolated onto $\Delta P = 0.2$ dbar steps while the final data are averaged within $\Delta P = 2.0$ dbar intervals according to stage 5. A linear trend was removed by least square trend elimination and a Hanning data window was used. There is a significant change in spectral shape and level in salinity from raw data and processed data. The raw data show a white noise in the spectra for a vertical scale smaller than 1 dbar indicating the effective vertical resolution. This flattening of spectra is shifted up to 5 and 6 dbar for the processed data. There is no significant difference between MAXDIF and MEDFIL (MEDDIF coincides with MAXDIF) while at higher wavenumbers the energy density of STADEV compared to that of MAXDIF is almost one order of magnitude smaller. This indicates a loss of information.

References

Baker, Jr., D.J. (1981) "Ocean Instruments and Experiment Design" in "Evolution of Physical Oceanography" edited by B. Warren and C. Wunsch, p. 396-433.

Biermann, P.H., P. Diekmann and B.S. Fedders (1976) "Die Telemetrie der Kieler Multisonde" presented in INTEROCEAN '76, Ref. No. IO 76-403, Düsseldorf, F.R.G., 15.-19. June 1976.

Dantzler, H.L. (1974) "Dynamic Salinity Calibrations of Continuous Salinity/Temperature/Depth Data". Deep-Sea Res., Vol. 21, p. 675-682.

Fahrbach, E., W. Krauss, J. Meincke and A. Sy (1983) "Nordostatlantik '81 - Data report". Ber. Inst. Meereskunde Kiel, F.R.G. (in press).

Fahrbach, E., W. Krauss, J. Meincke and A. Sy (1983) "Nordostatlantik '82 - Data report". Ber. Inst. Meereskunde Kiel, F.R.G. (in preparation).

Fofonoff, N.P. and R.C. Millard "Algorithms for Oceanographic Computations". Submitted to Unesco Technical Papers in Marine Science.

Fofonoff, N.P., S.P. Hayes and R.C. Millard (1974) "WHOI/ Brown CTD Microprofiler: Methods of Calibration and Data Handling" Woods Hole Oceanographic Institution, WHOI-74-48, Woods Hole MA. U.S.A., 66 pp (unpublished manuscript).

Gregg, M.C., J.C. Schedvin, W.C. Hess and T.B. Meagher (1982) "Dynamic Response Calibration of the Neil Brown Conductivity Cell". J. Phys. Oc., Vol. 12, p. 720-742.

Grose, P.L. (1981) "CTD Data Processing". Center for Environmental Assessment Services, NOAA, Washington DC, 20235, U.S.A. (unpublished manuscript).

Horne, E.P.W. and J.M. Toole (1980) "Sensor Response Mismatches and Lag Correction Techniques for Temperature-Salinity Profilers".
J. Phys. Oc., Vol. 10, p. 1122-1130.

Hurst, W.S. (1975) "Note on the Measurement of the Response of Oceanographic Temperature Sensors".
J. Geophys. Res., Vol. 80, p. 2663-2666.

Joyce, T.M. (1976) "Large-Scale Variations in Small-Scale Temperature/Salinity Finestructure in the Main Thermocline of the Northwest Atlantic".
Deep-Sea Res., Vol. 23, p. 1175-1186.

Justusson, B.J. (1981) "Median Filtering: Statistical Properties" in "Two-Dimensional Digital Signal Processing II".
Springer-Verlag Berlin-Heidelberg-New York, p. 161-196.

Kroebel, W., P. Diehl, L. Ginzkey, K.-H. Mahrt, J. Rathlev, R. Siara and Th. Schulz (1976) "Die Kieler Multisonde der Jahre 1975/76, ihre Sensoren, Parameter mit Ergeb nissen von Datenaufnahmen und Perspektiven für ihre Auswertungen" presented in INTEROCEAN '76, Ref. No. IO 76-402, Düsseldorf, F.R.G., 15.-19. June 1979, p. 1034-1046.

Langhof, H.-J. and W. Zenk (1982) "Die Gallium-Tripelpunkt-zelle als Kontrollinstrument für hochgenaue ozeano graphische Thermometer".
Technischer Bericht, Inst. für Meereskunde Kiel, F.R.G., 17 pp. (unpublished manuscript).

Meerestechnik-Elektronik GmbH, "Handbuch I, II für die Multisonde".

Perkin, R.G. and E.L. Lewis (1982) "Design of CTD Observational Programmes in Relation to Sensor Time Constants and Sampling Frequencies".
Can. Techn. Rep. Hydrogr. Ocean Sci., No. 7, 47 pp.

Peters, H. (1976) "GATE - CTD data Measured on the F.R.G.
    Ships Shipboard Operations-Calibrations-Editing".
    Technical Report, Ber. Inst. Meereskunde Kiel, F.R.G.,
    27 pp.

Rathlev, J. (1981) "DT7 - Data Transfer Programme Rev. 7.5,
    User's Manual". Institut für Angewandte Physik, Kiel,
    F.R.G. (unpublished manuscript).

Roden, G.J. and J.D. Irish (1975) "Electronic Digitization
    and Sensor Response Effects on Salinity Computation from
    CTD Field Measurements".
    J. Phys. Oc., Vol. 5, p. 195-199.

Scarlet, R.J. (1975) "A Data Processing Method for Salinity,
    Temperature, Depth Profiles".
    Deep-Sea Res., Vol. 22, p. 509-515.

Tukey, J.W. (1971) "Exploring Data Analysis".
    Addison-Wesley, Reading, MA., U.S.A.

Tyan, S.G. (1981) "Median Filtering: Deterministic Proper-
    ties" in "Two-Dimensional Digital Signal Processing II".
    Springer-Verlag Berlin-Heidelberg-New York, p. 197-217.

Unesco (1981) "Background papers and supporting data on the
    Practical Salinity Scale 1978".
    Unesco Technical Papers in Marine Science, No. 37,
    144 pp.

DATA FLOW DIAGRAM FOR THE CTD-SYSTEM „MULTISONDE"
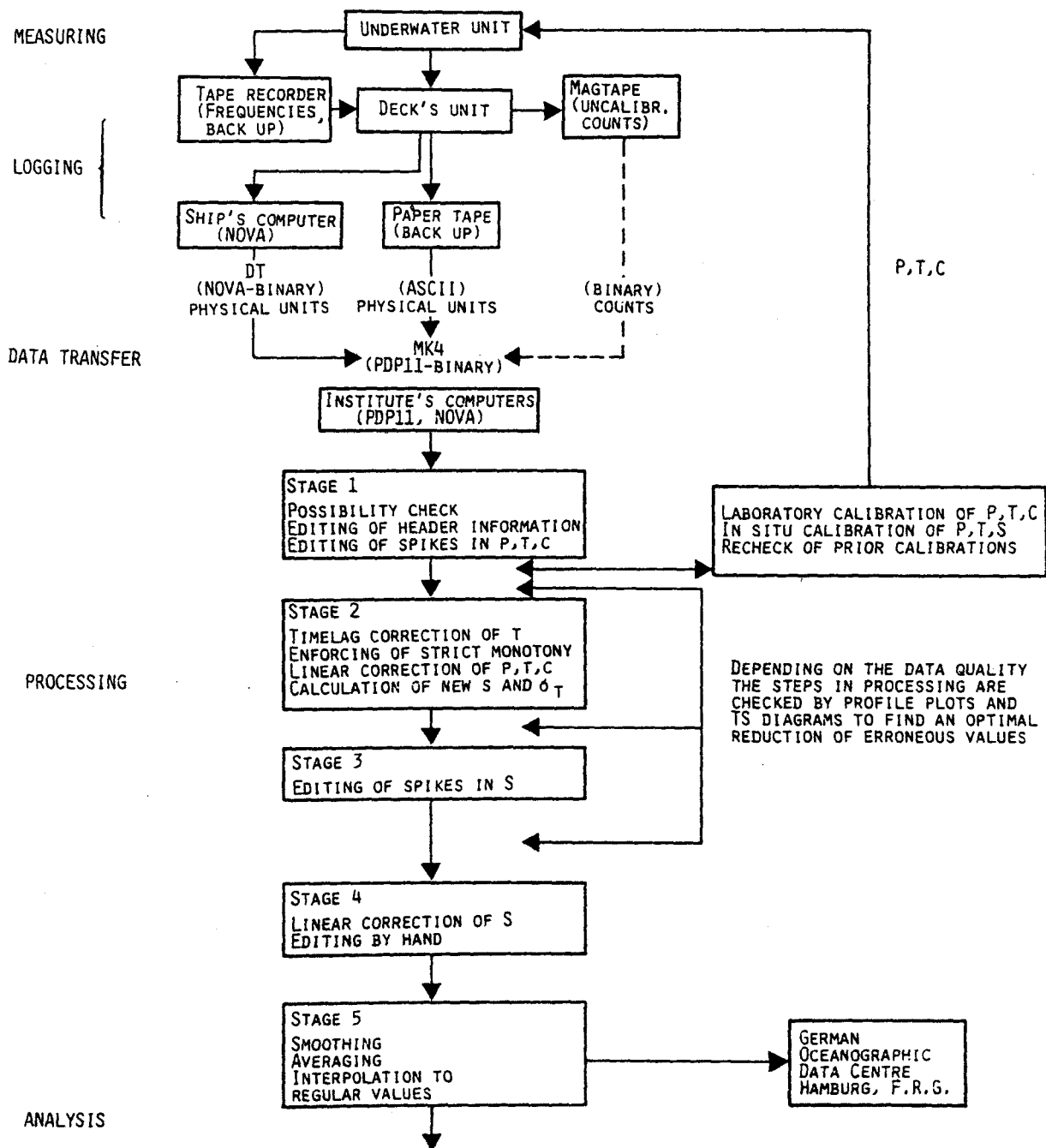
Fig. 1.: Flow diagram for acquisition and principles of processing of "Multisonde" CTD data.
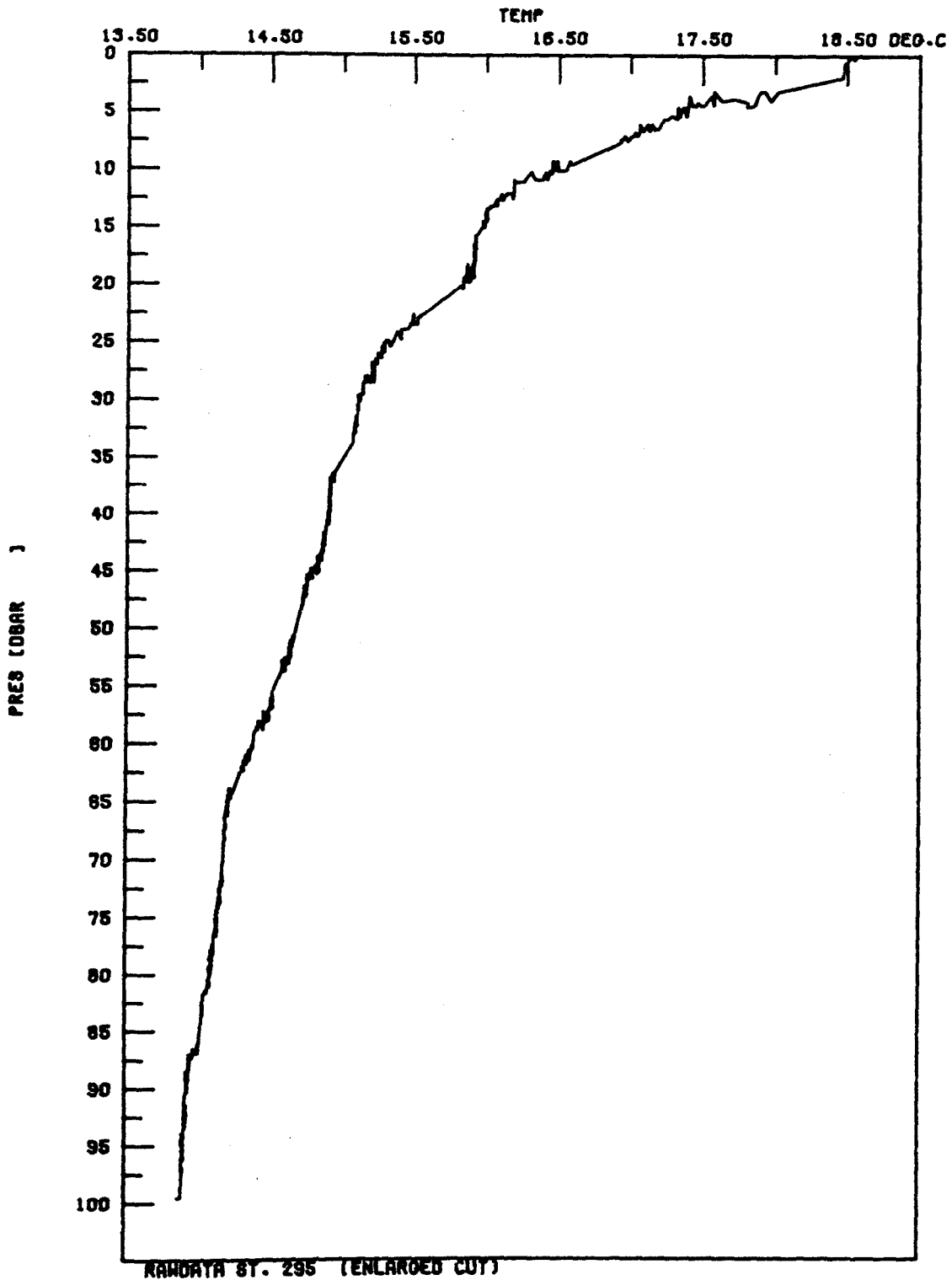
Fig. 2a-d: The influence of ship's roll on records of (a) temperature, (b) conductivity, (c) salinity and (d) reduction of salinity spikes after enforcing rigid monotonic sequence of pressure and calculation of new salinity.

RAWDATA ST. 295 (ENLARGED CUT)

Fig. 2b

Fig. 2c

ST. 295. RIGID MONOTONY, NEW SALINITY (ENLARGED CUT)
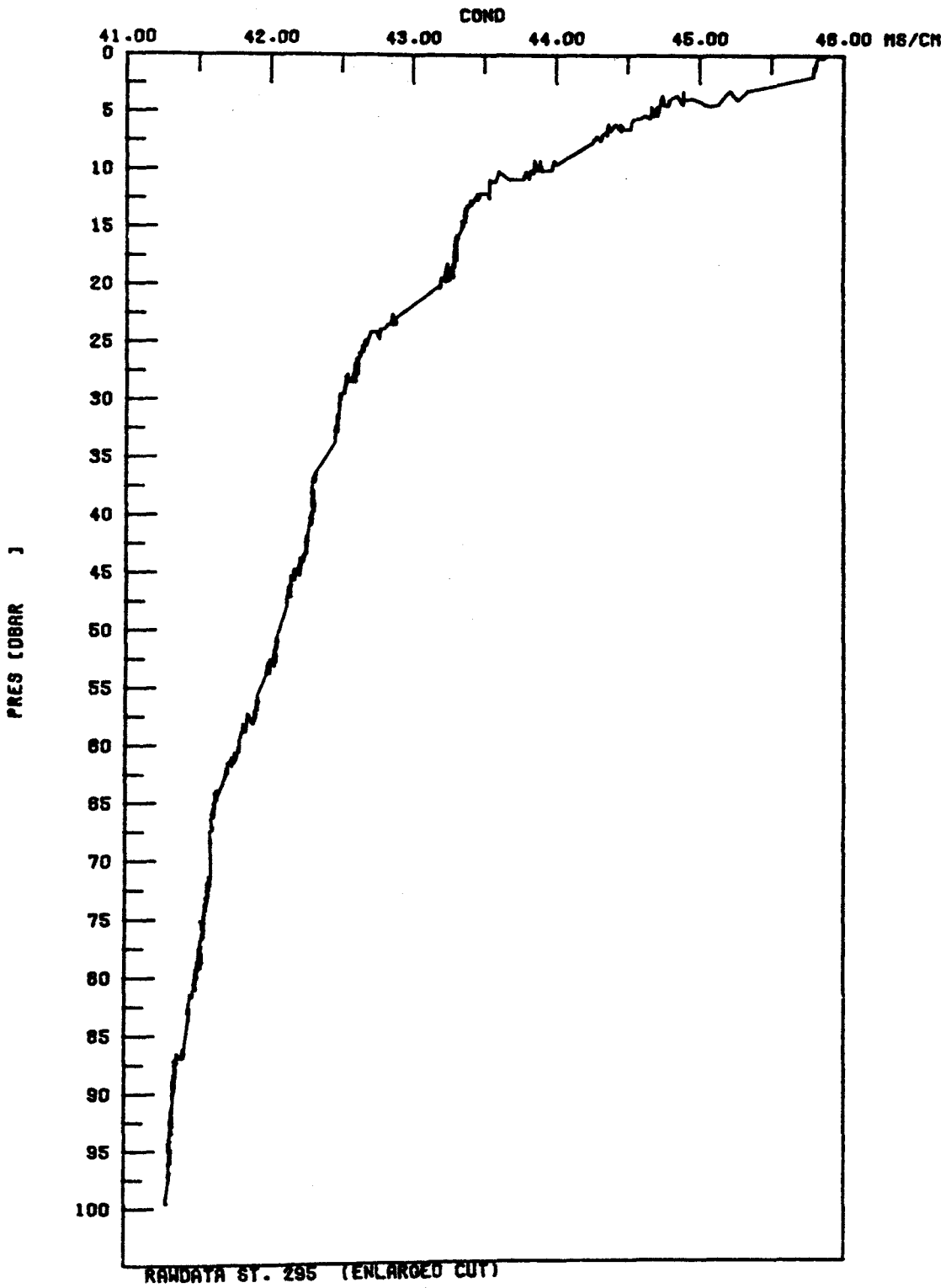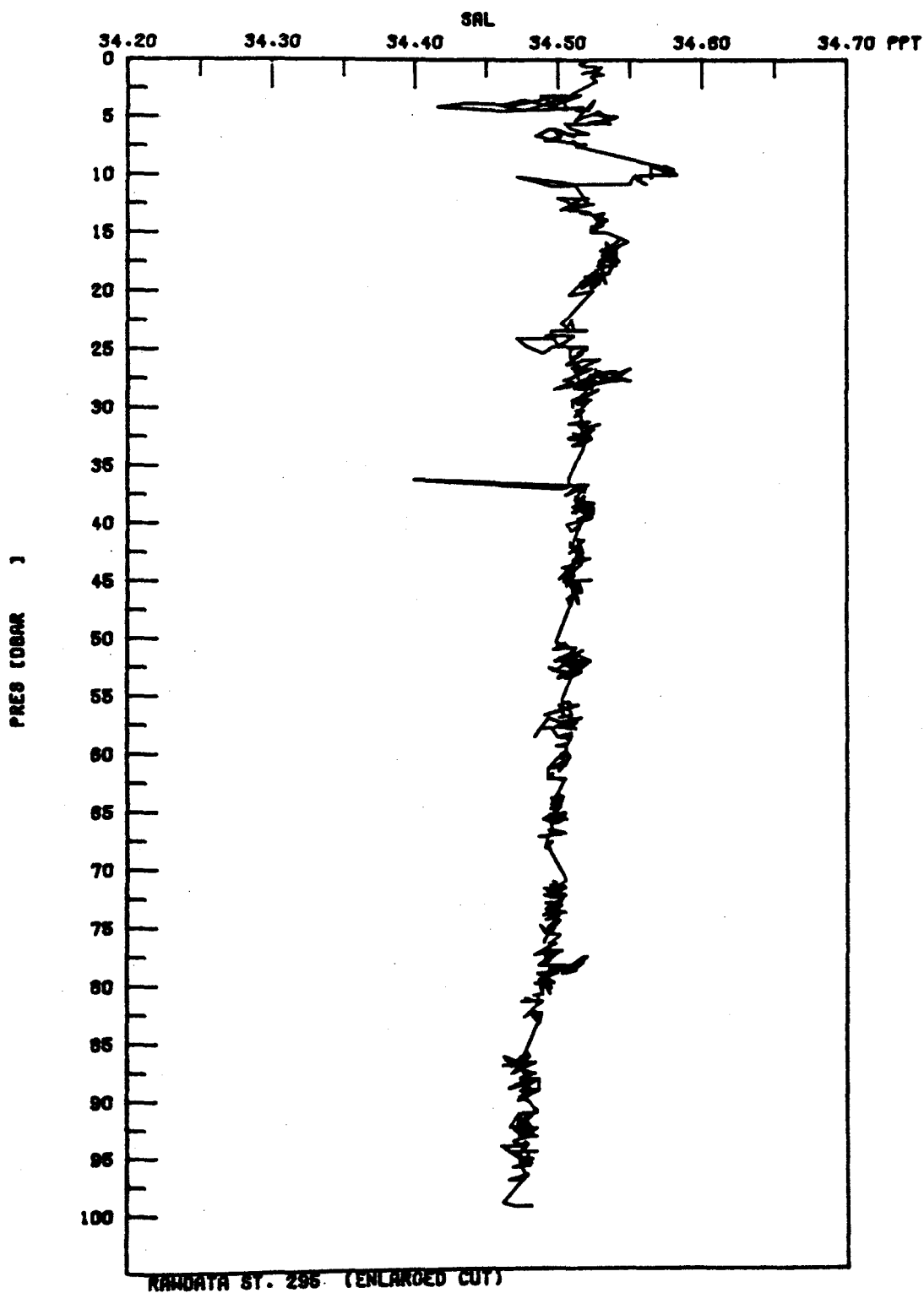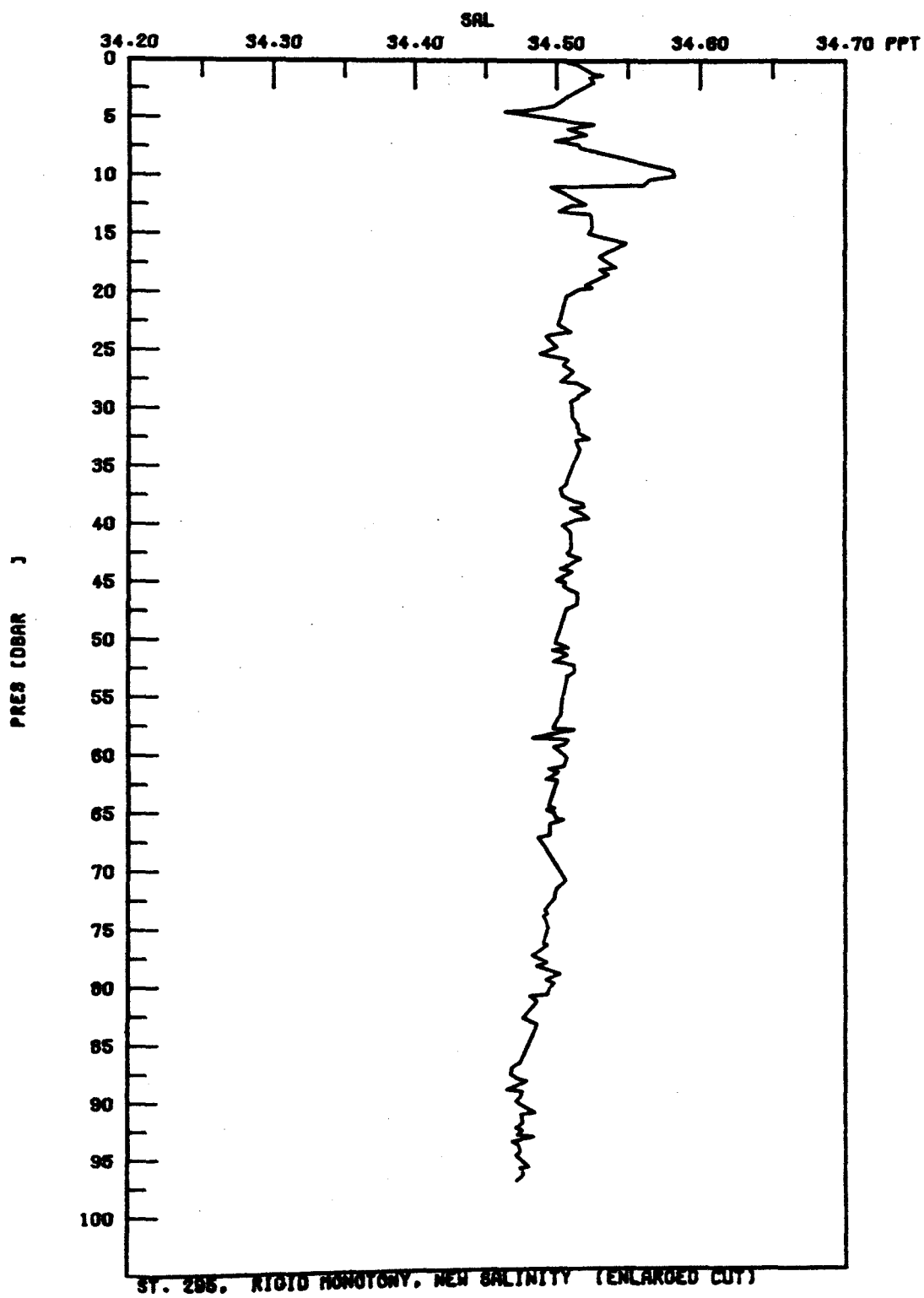
Fig. 2d
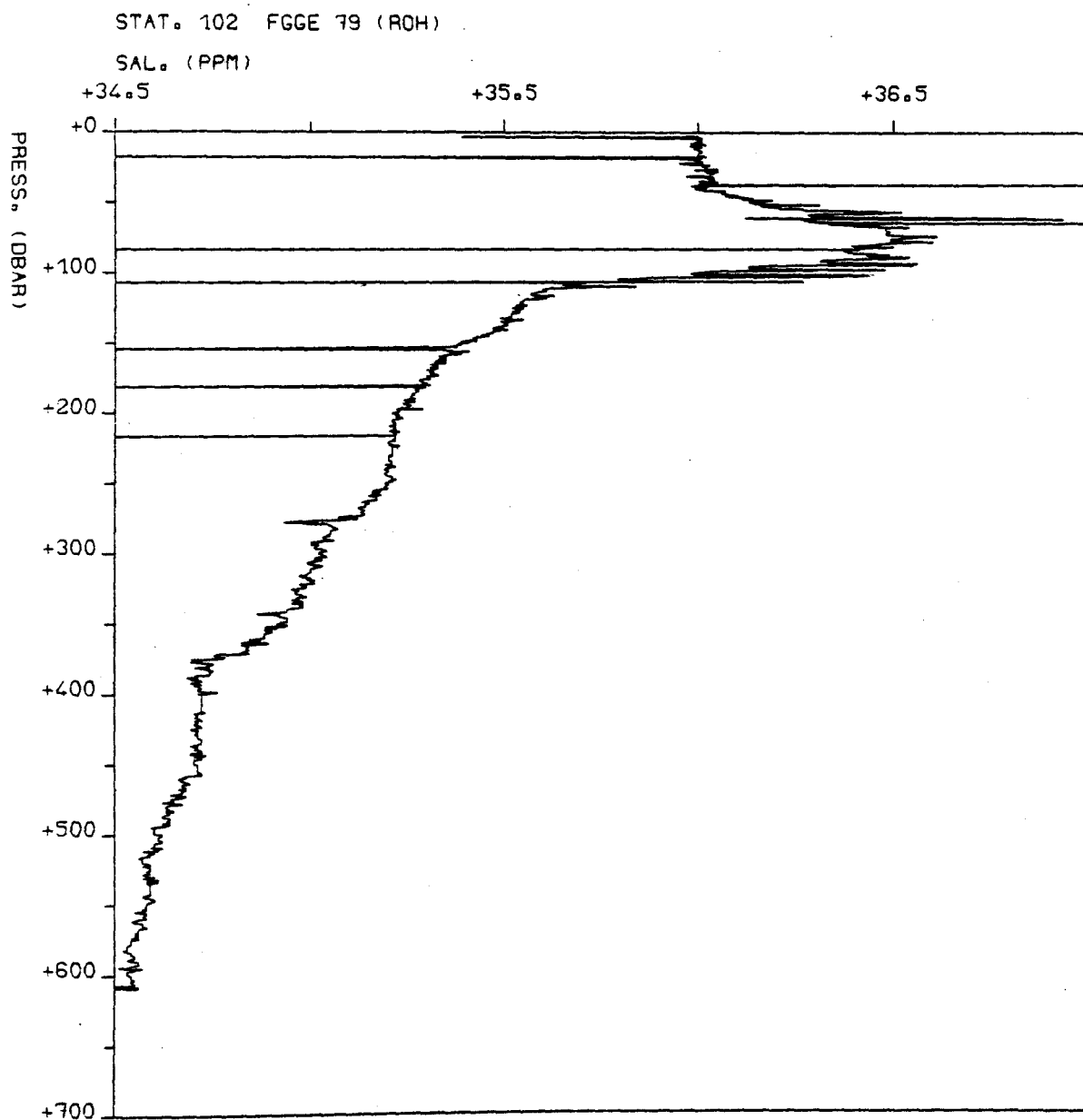
Fig. 3: Example of spikes in a salinity profile which are difficult to remove sufficiently. The signal is contaminated by slip ring noise. The data are collected with the "Bathysonde" during FGGE '79.
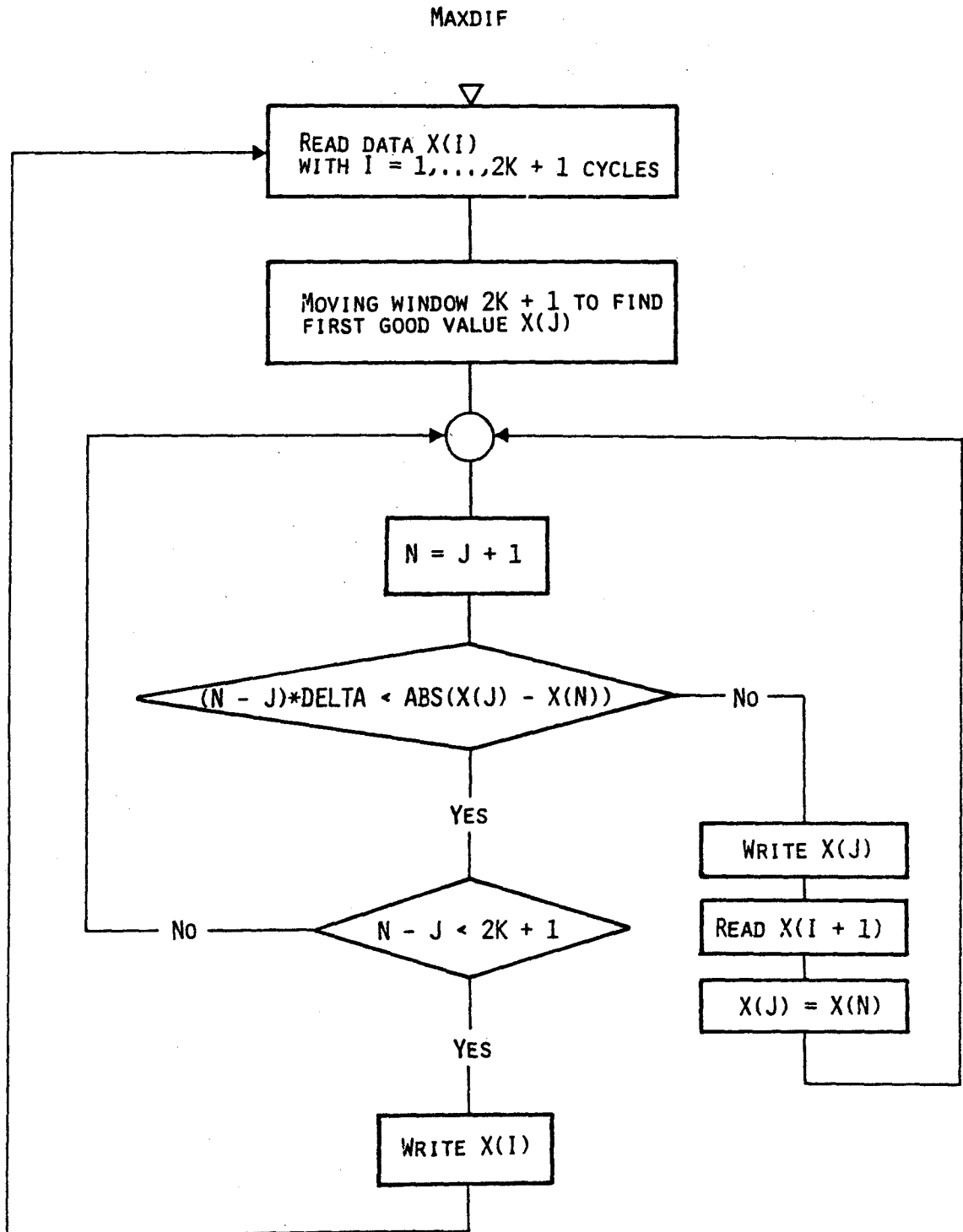
MAXDIF

FREE PARAMETER: 1. WINDOW LENGTH Q = 2K + 1

2. DELTA = DESIRED MAXIMUM DIFFERENCE
   BETWEEN TWO ADJACENT VALUES

Fig. 4:   Flow diagram for editing program MAXDIF.

STADEV
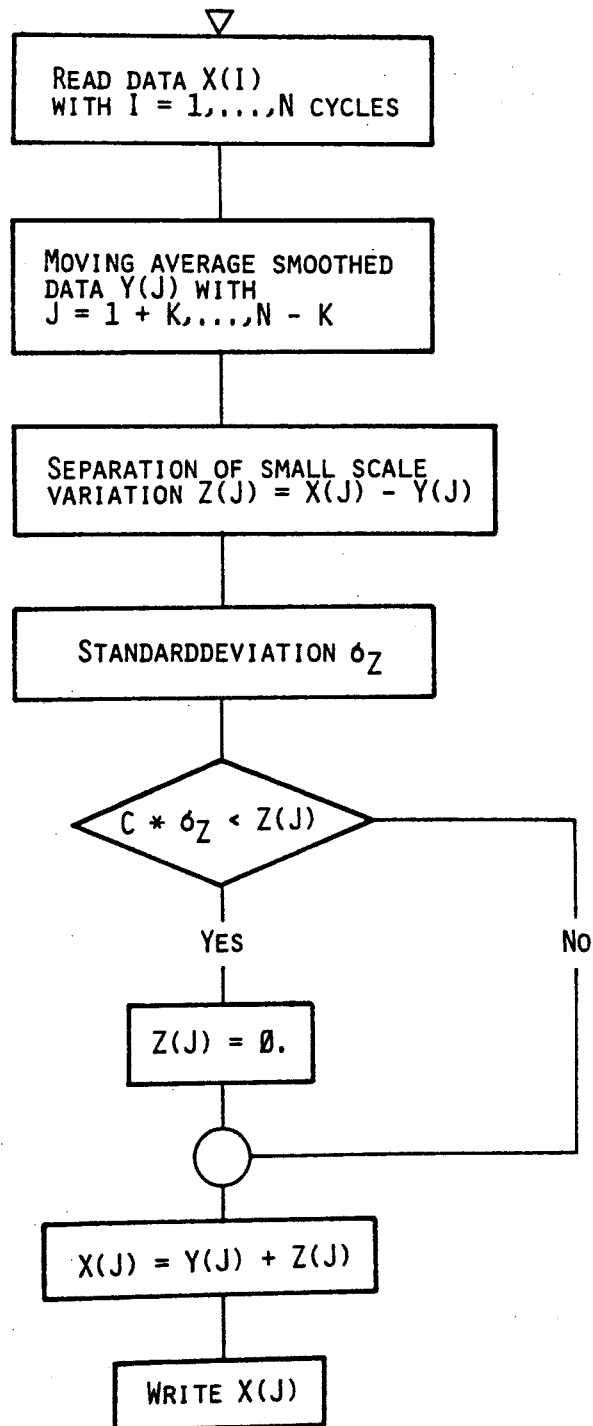
FREE PARAMETER: 1. WINDOW LENGTH $Q = 2K + 1$ FOR MOVING AVERAGE

2. FACTOR C

Fig. 5: Flow diagram for editing program STADEV.
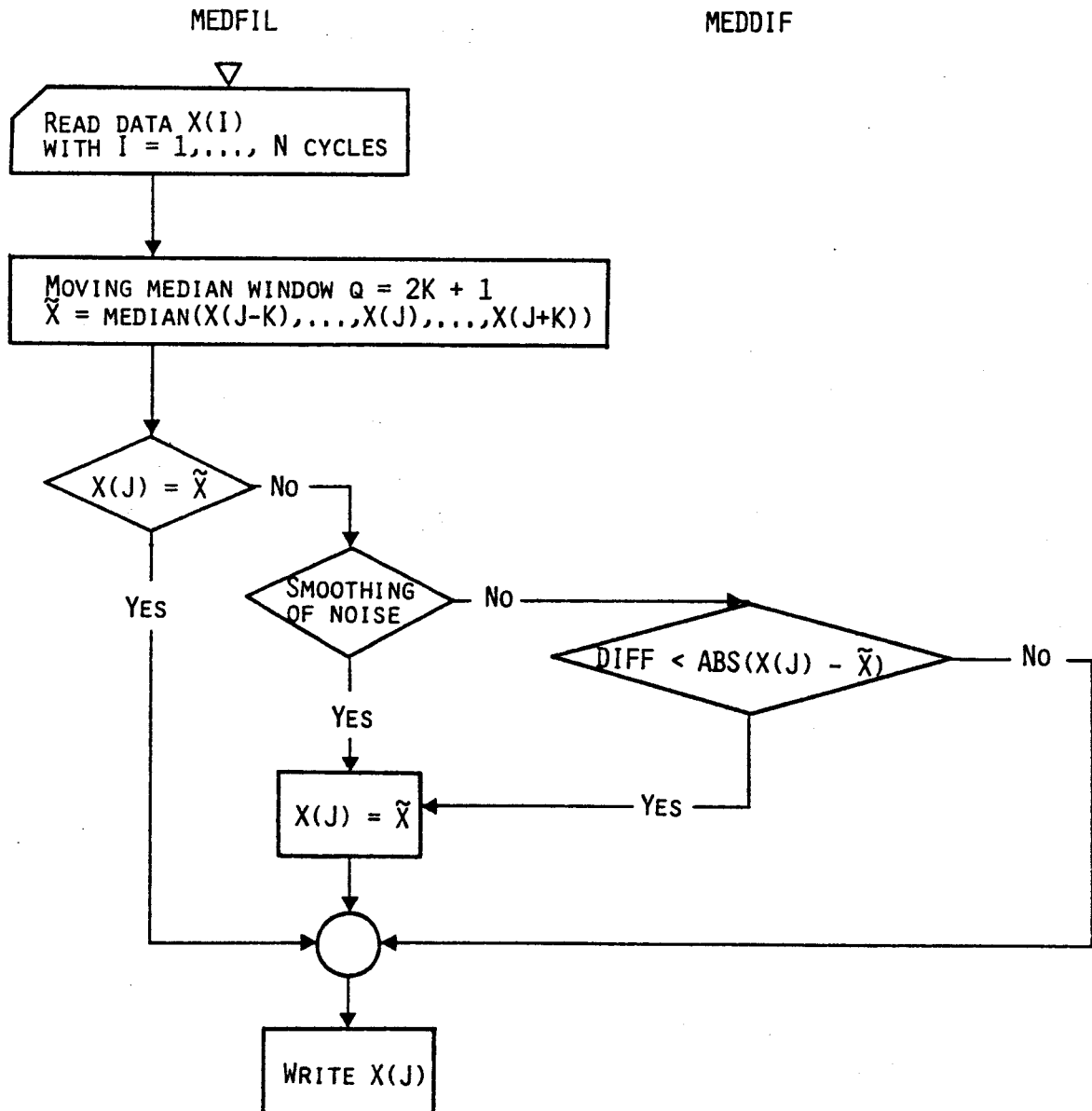
MEDFIL   MEDDIF

FREE PARAMETER: 1. WINDOW LENGTH Q = 2K + 1
                2. DIFF = DESIRED MAXIMUM DIFFERENCE
                         BETWEEN CONSIDERED VALUE AND
                         MEDIAN OF THIS WINDOW
                         (NO SMOOTHING OF NOISE)

Fig. 6:   Flow diagram for editing programs MEDFIL and
             MEDDIF.
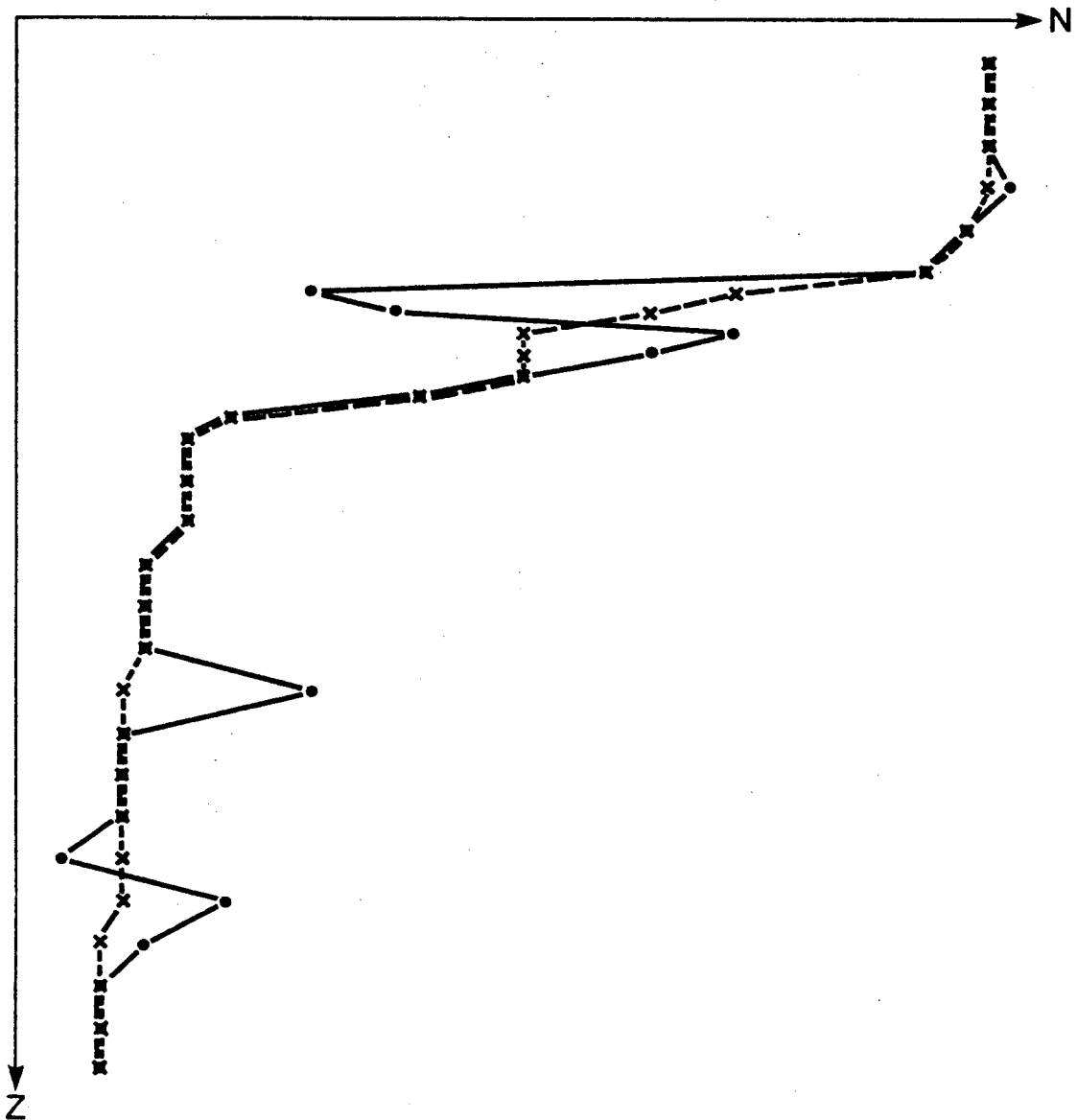
Fig. 7:   Illustration of principle properties for a five-
point median filter. The solid line is before
median filtering. The dashed line is after
median filtering.
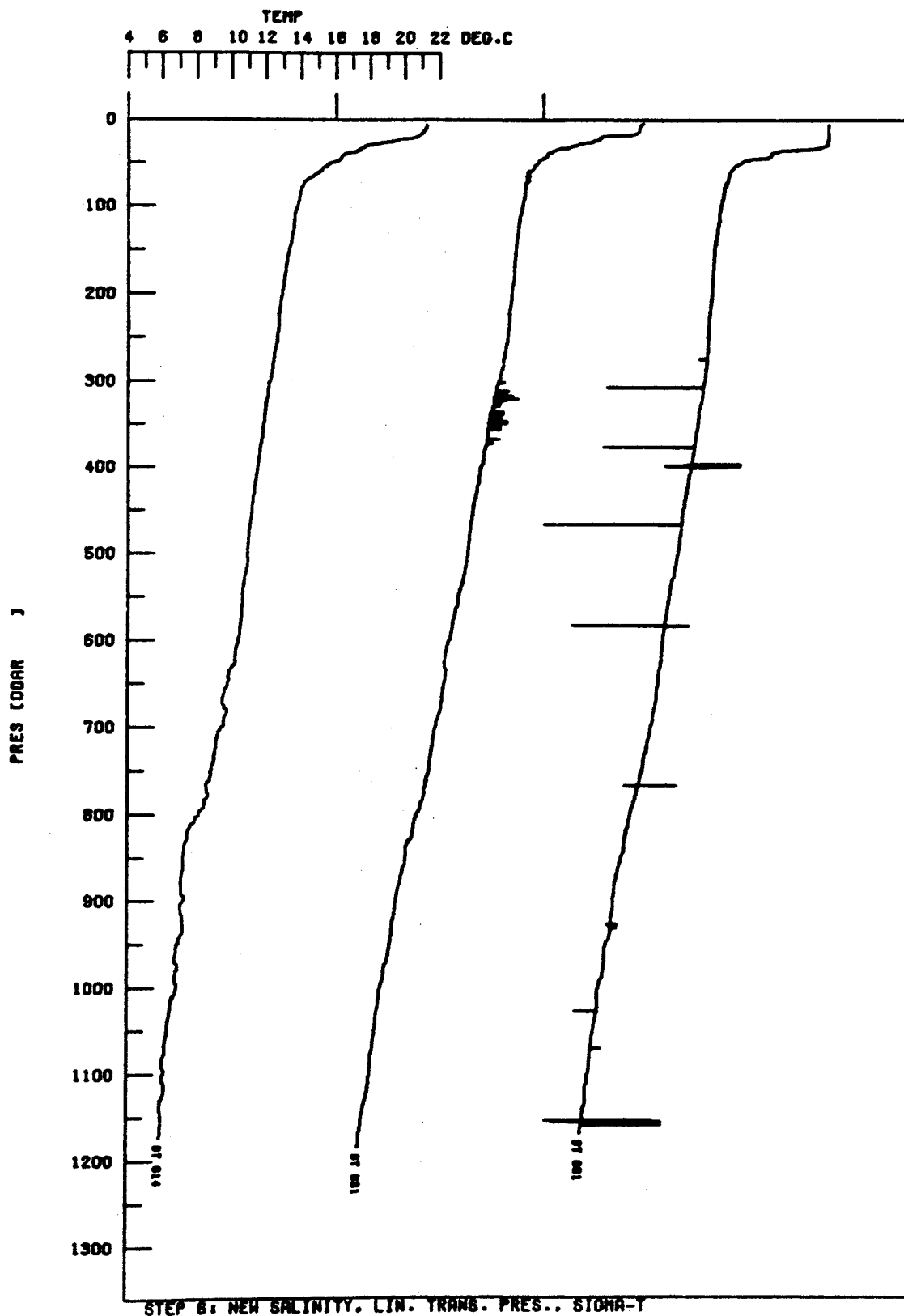
Fig. 8a-d: Raw data of the three chosen profiles St. 614, St. 631 and St. 661 for (a) temperature, (b) conductivity, (c) salinity and (d) density.
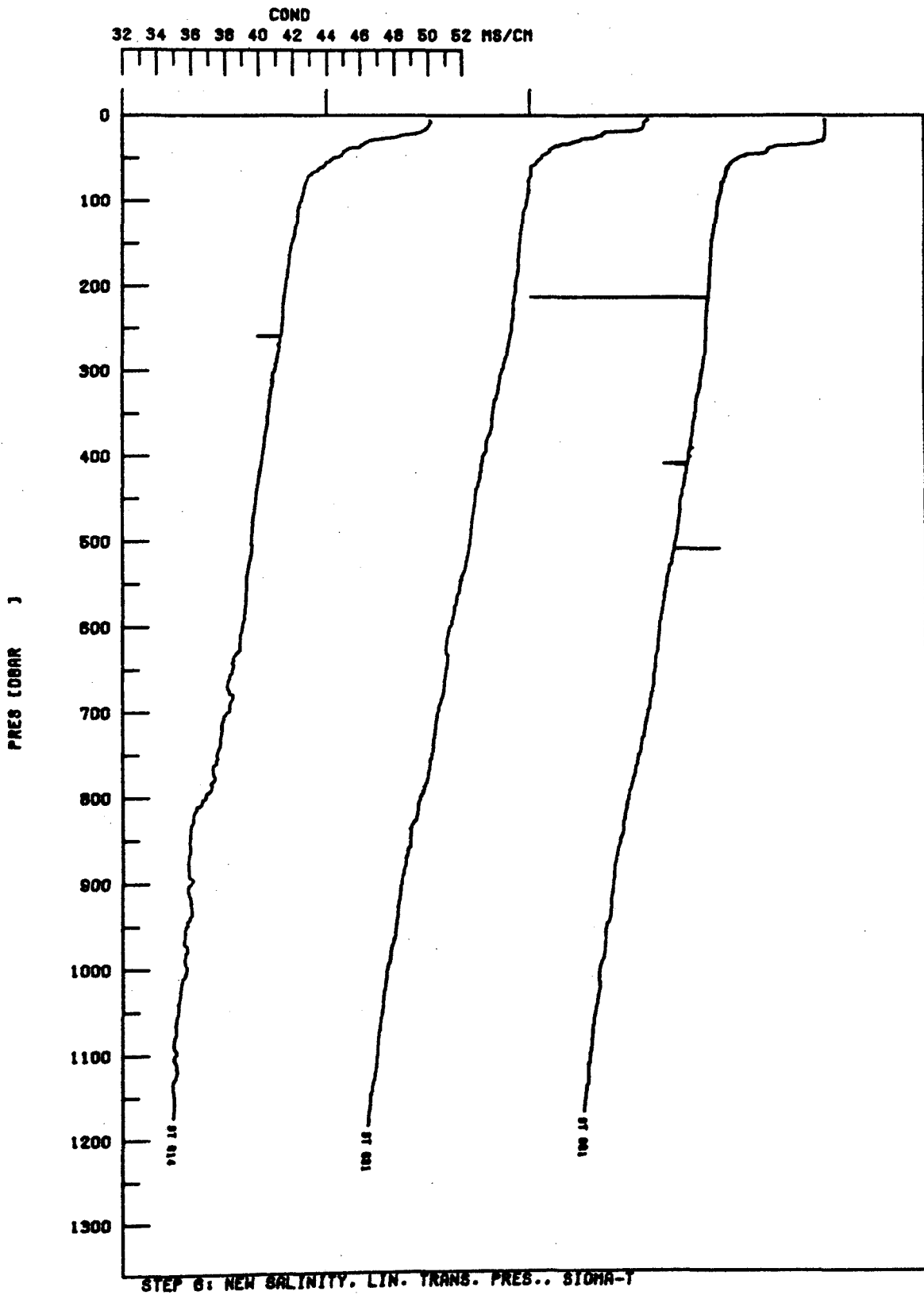
Fig. 8b
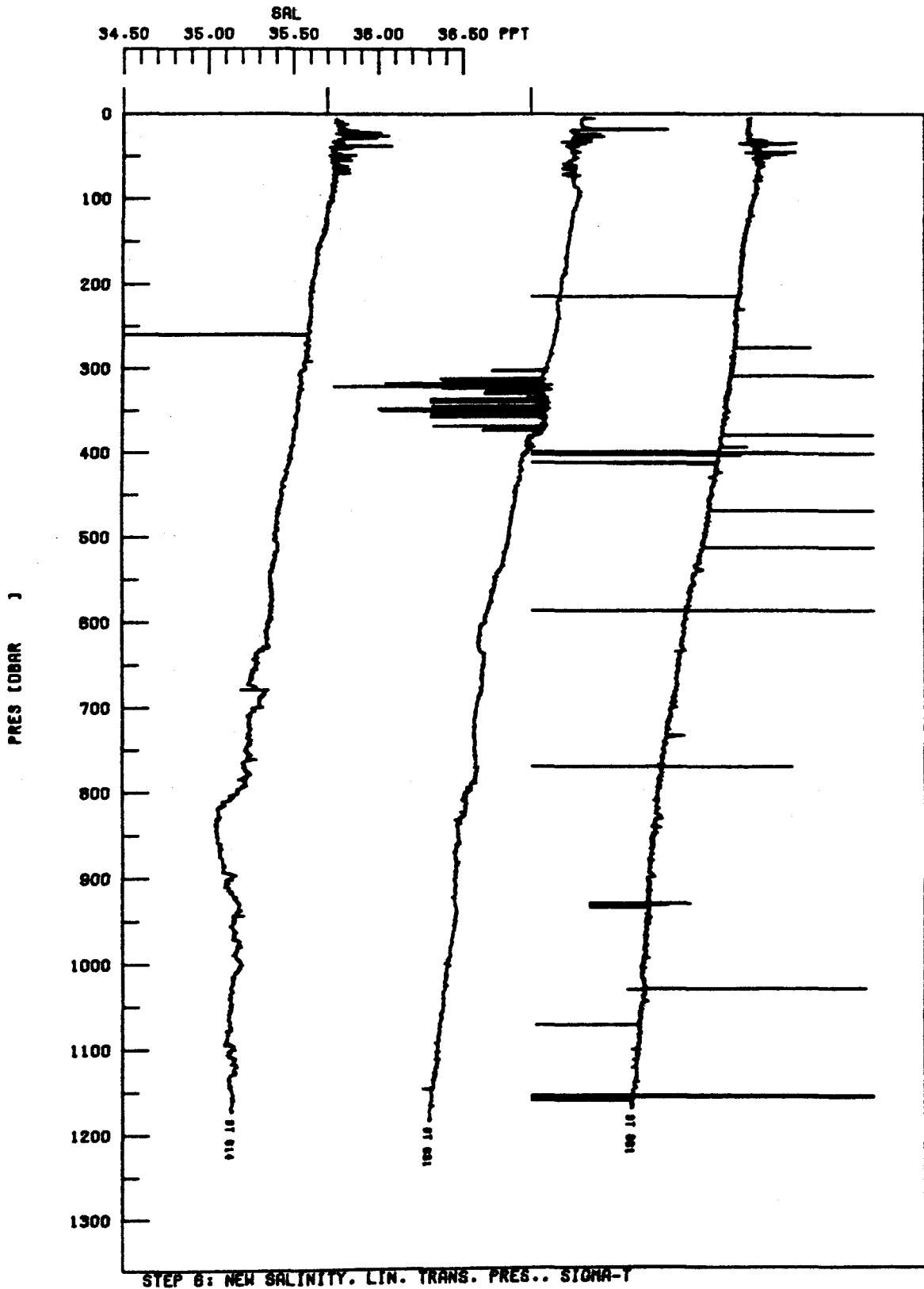
STEP 6: NEW SALINITY. LIN. TRANS. PRES.. SIGMA-T

Fig. 8c
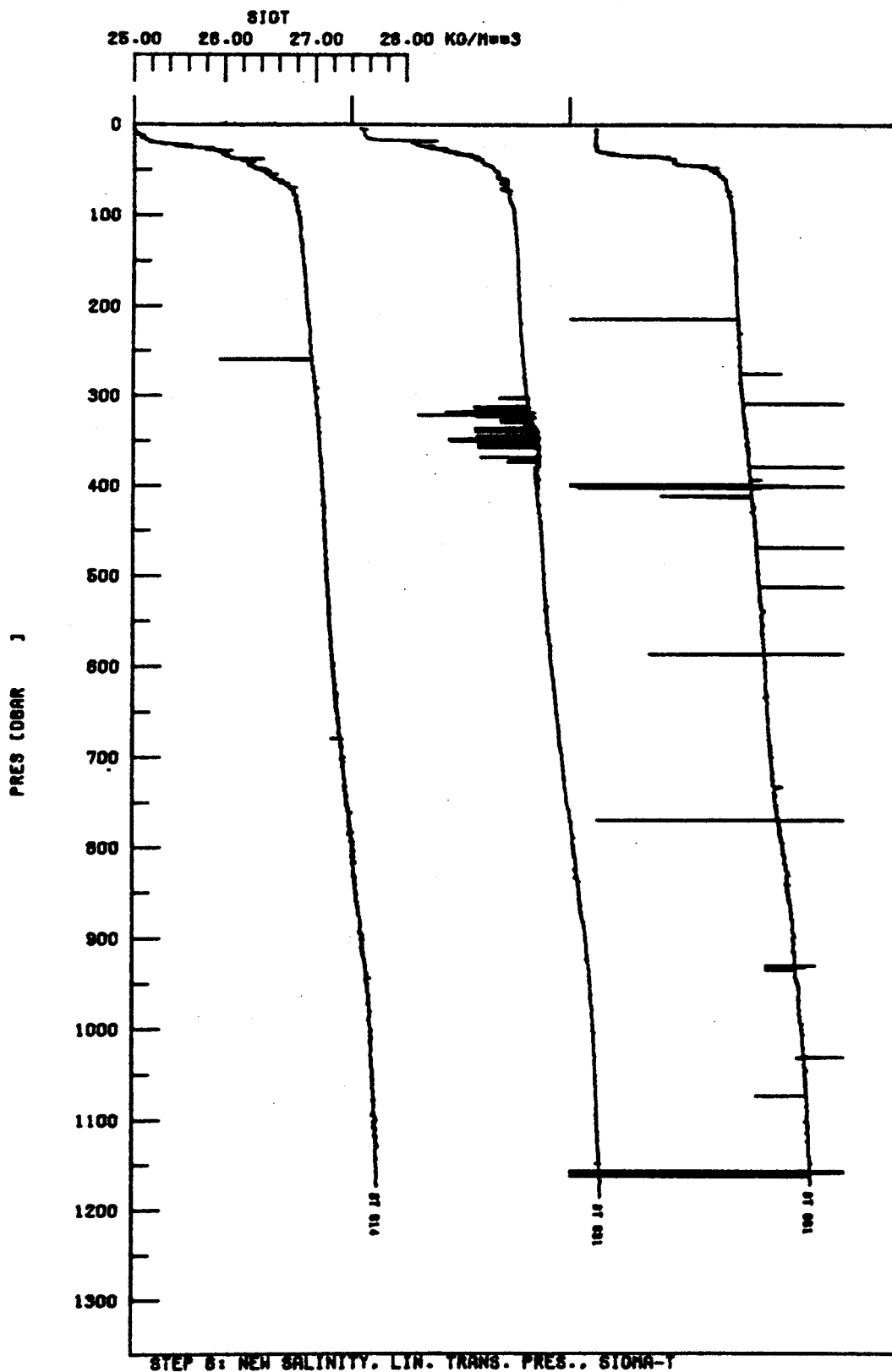
Fig. 8d
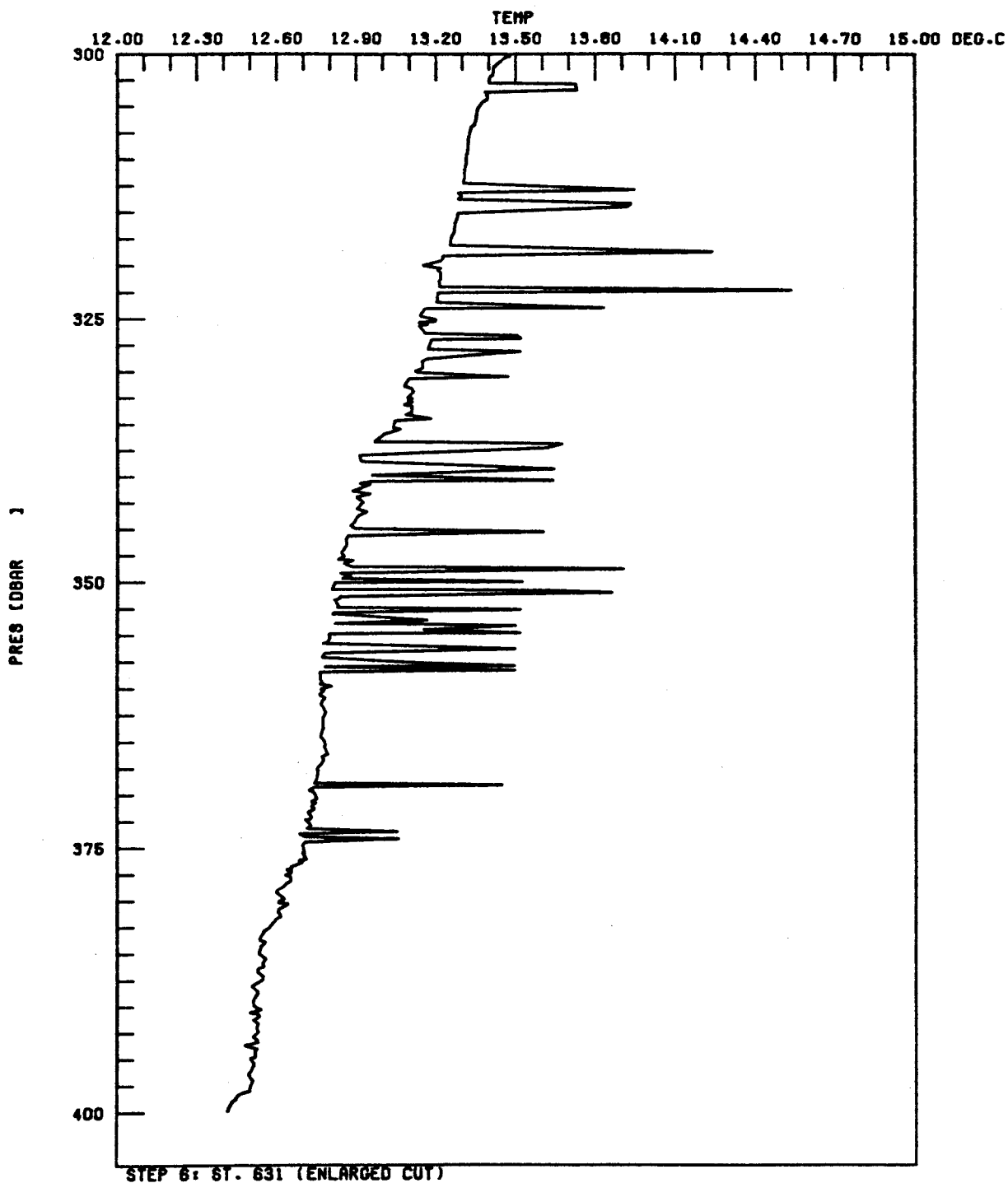
Fig. 9a: Enlarged cut of the temperature record of the spiky part of St. 631 shown in fig. 8a.
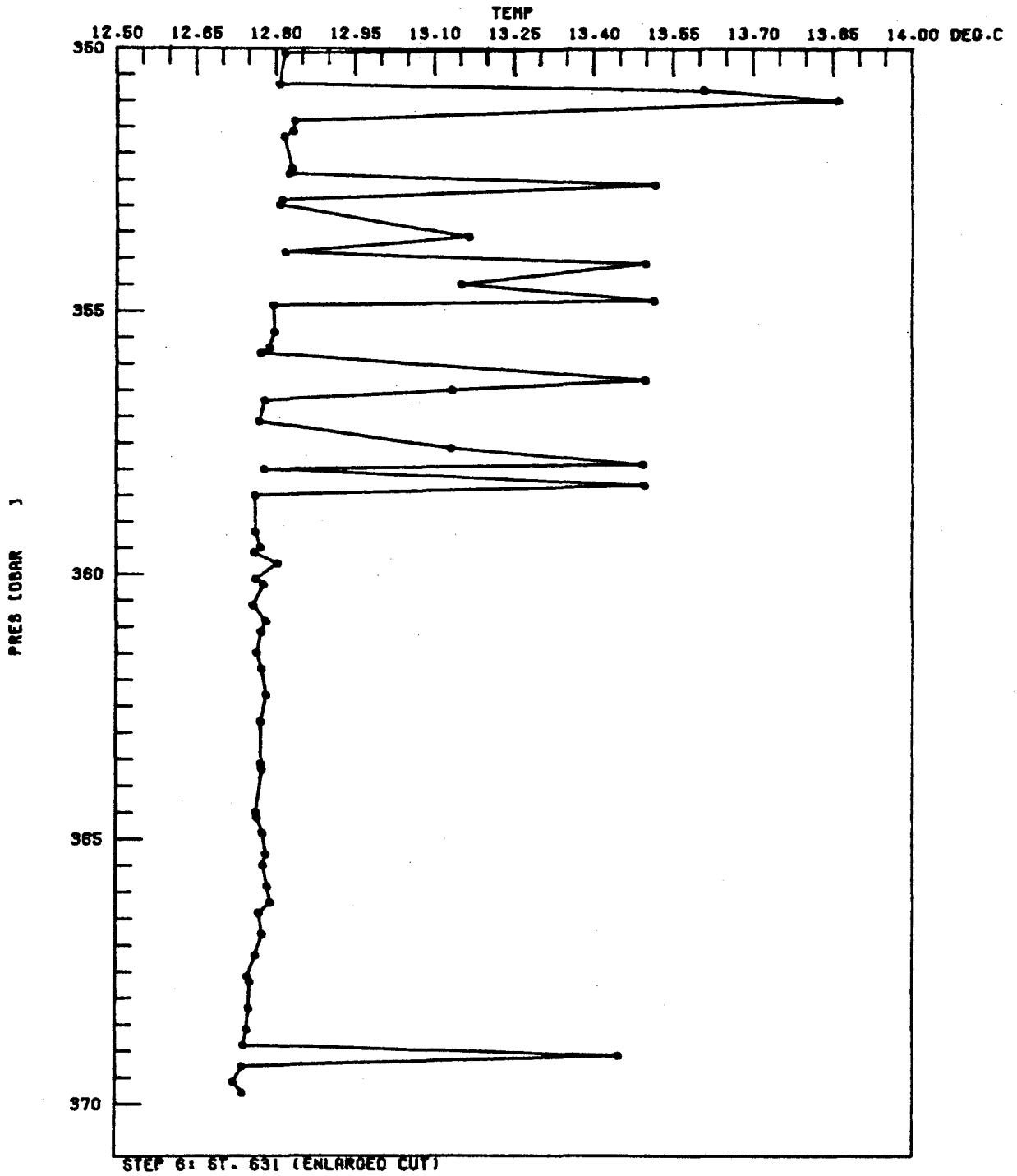
Fig. 9b: Same as fig. 9a, but with data point resolution.
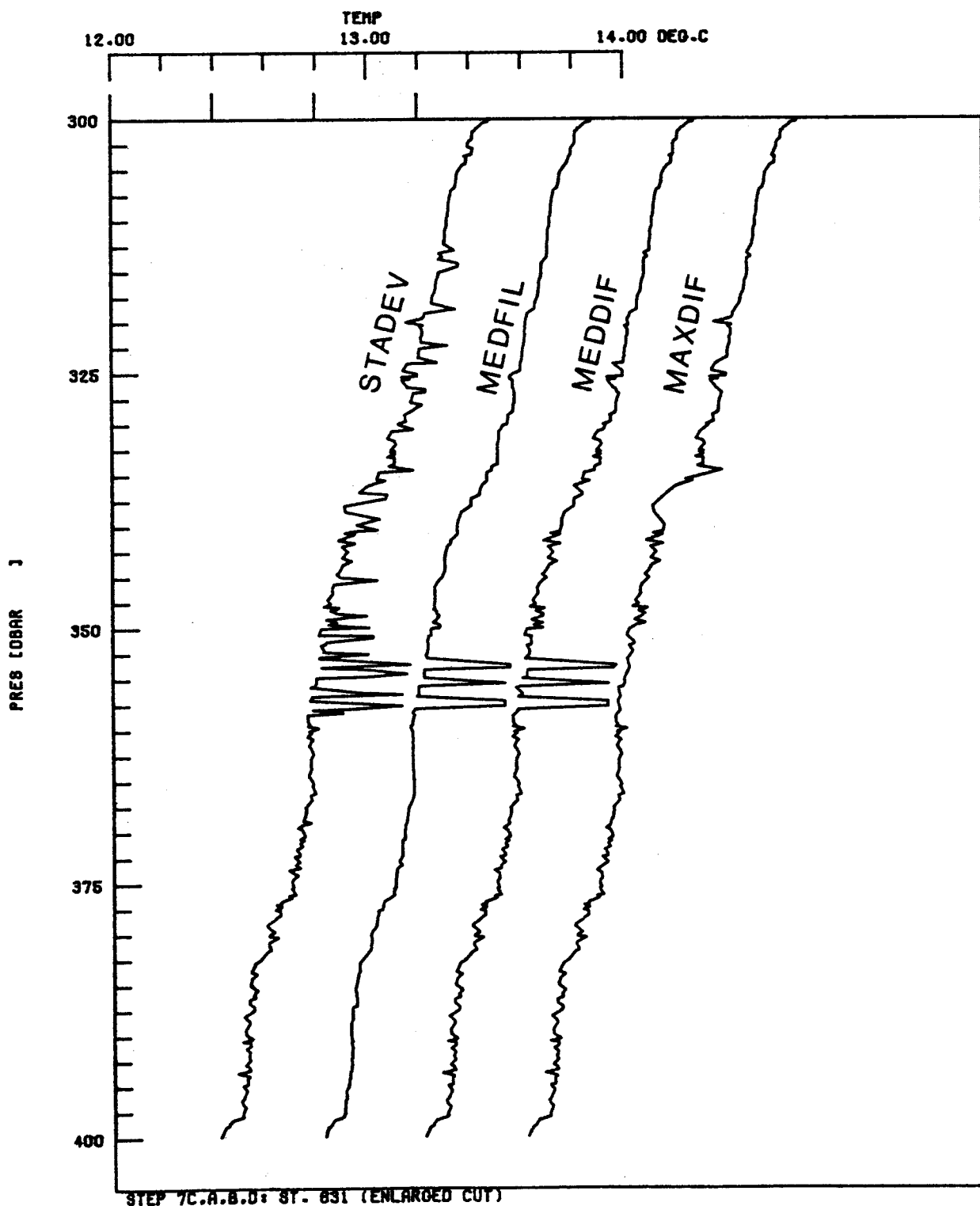
Fig. 10: Comparison of results obtained by the four editing techniques after the first run.
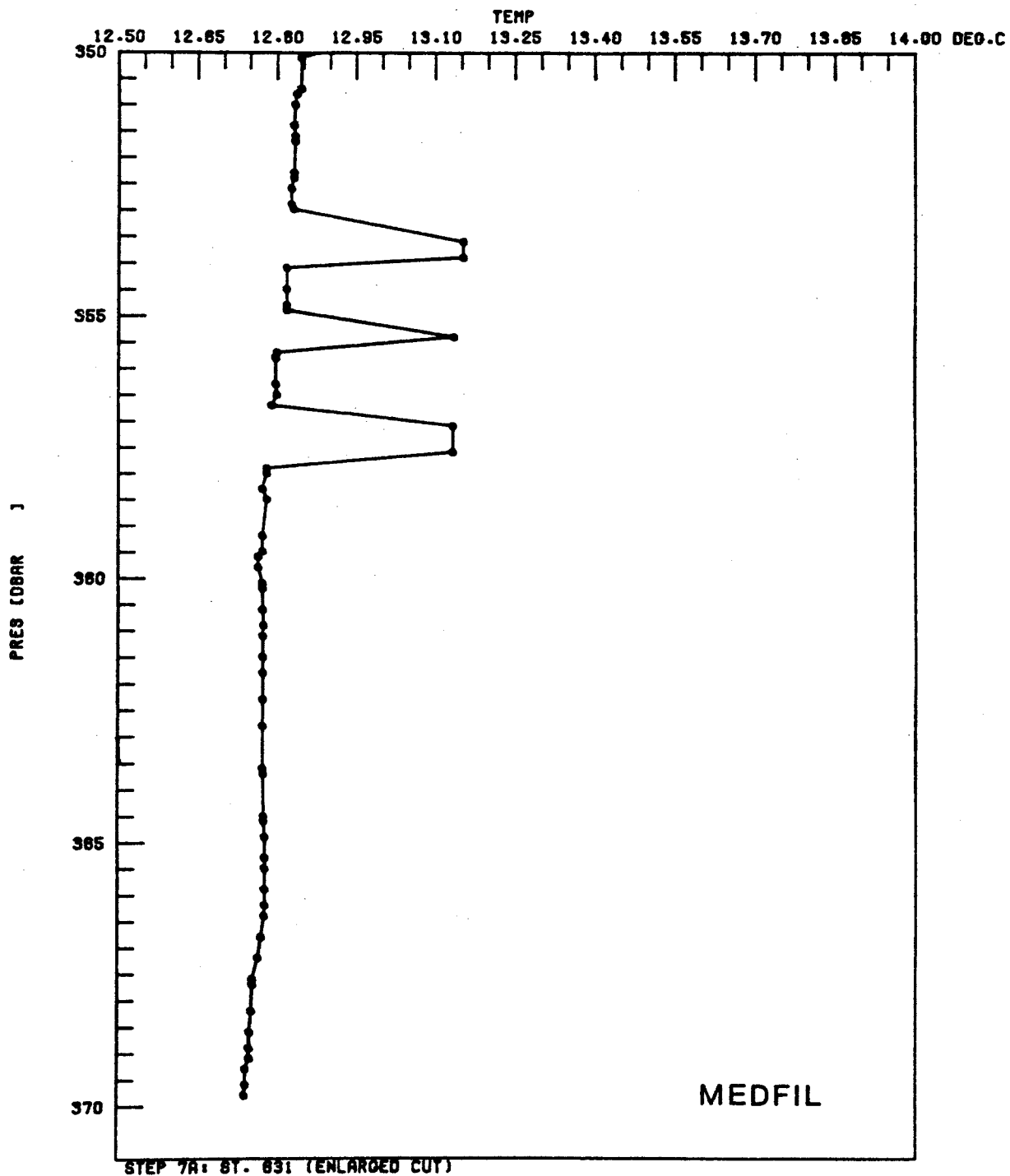
Fig. 11a-d: Same as fig. 10, but with data point resolution
for STADEV, MEDFIL, MEDDIF and MAXDIF.
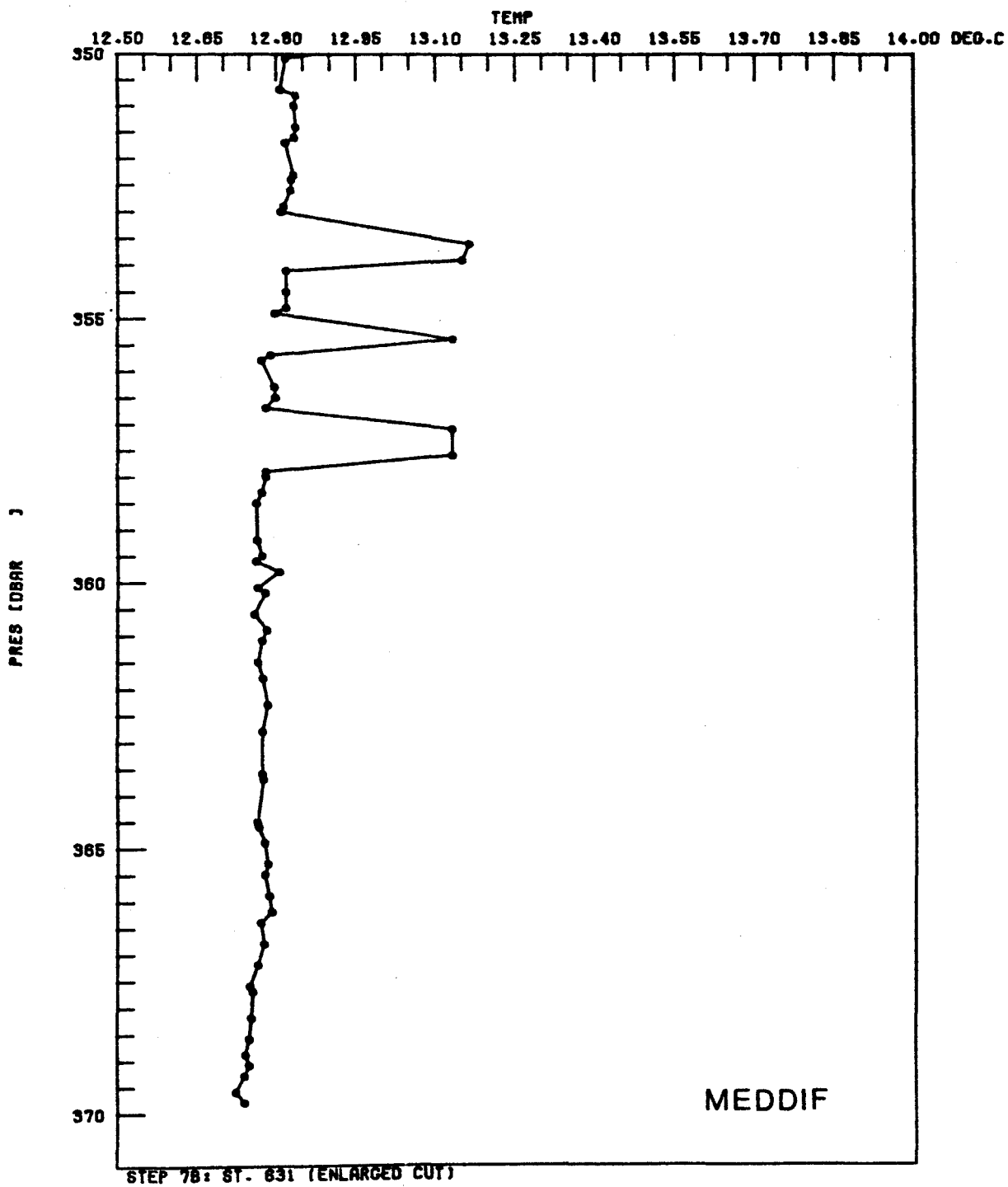
Fig. 11b
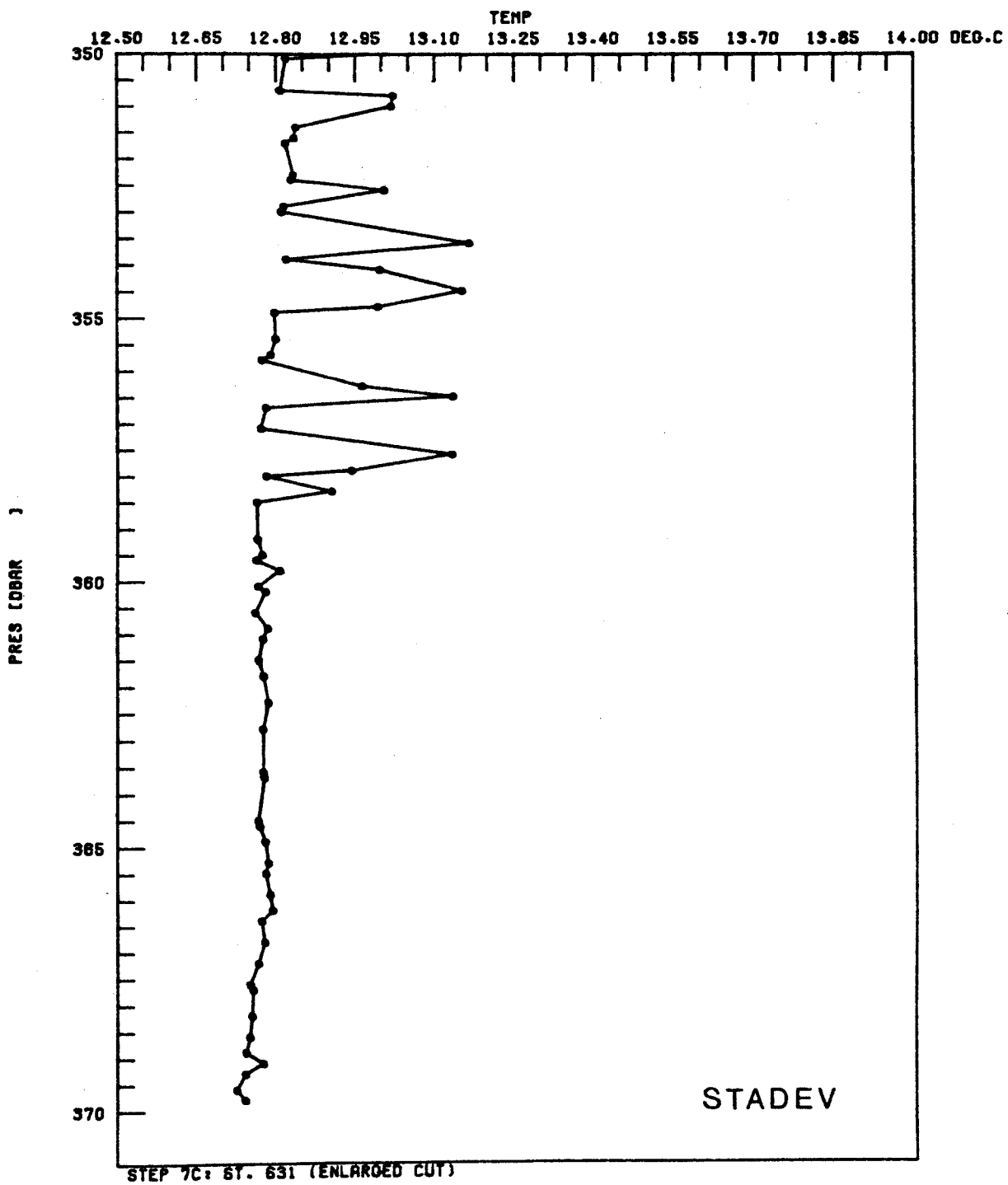
Fig. 11c
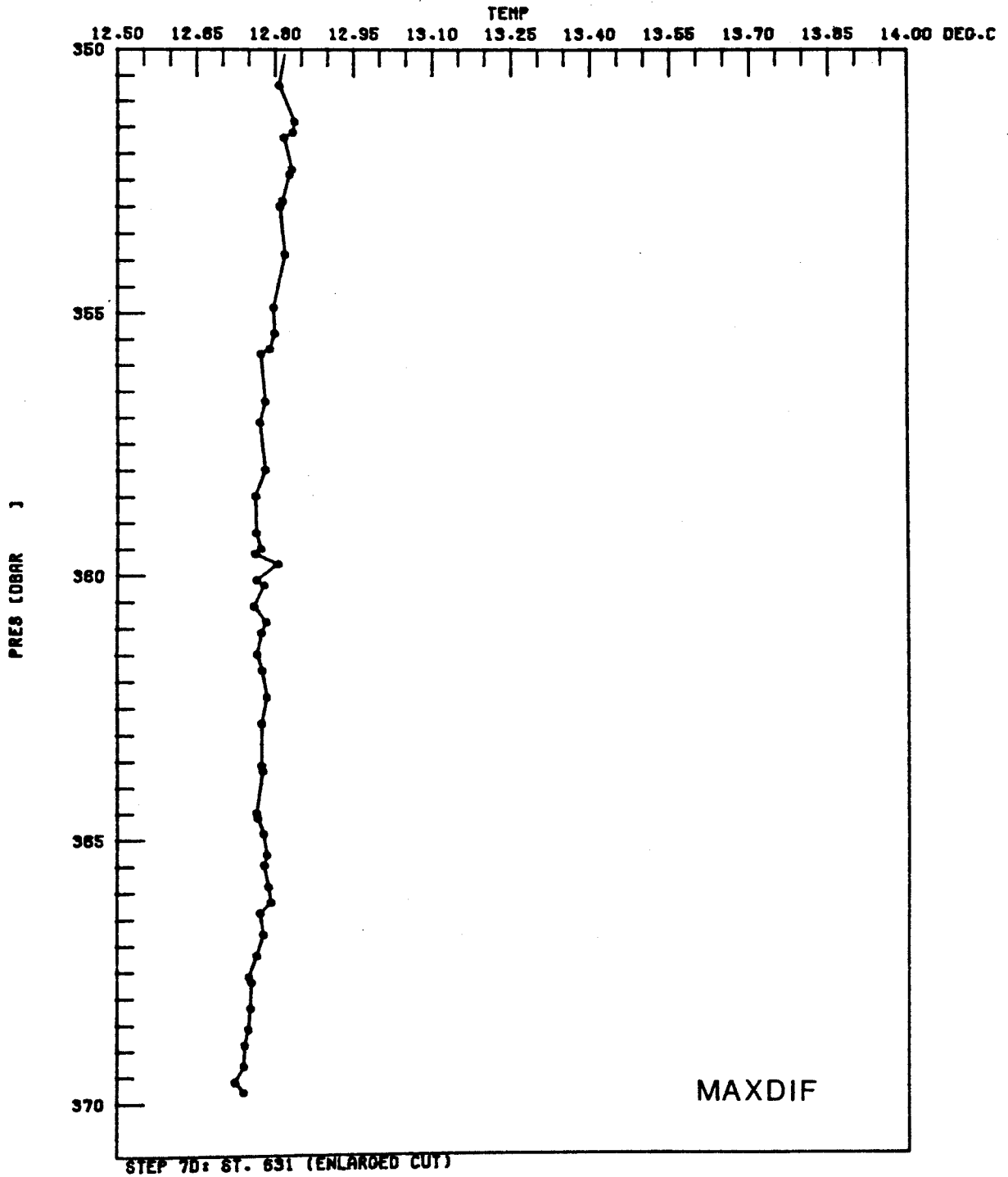
Fig. 11d
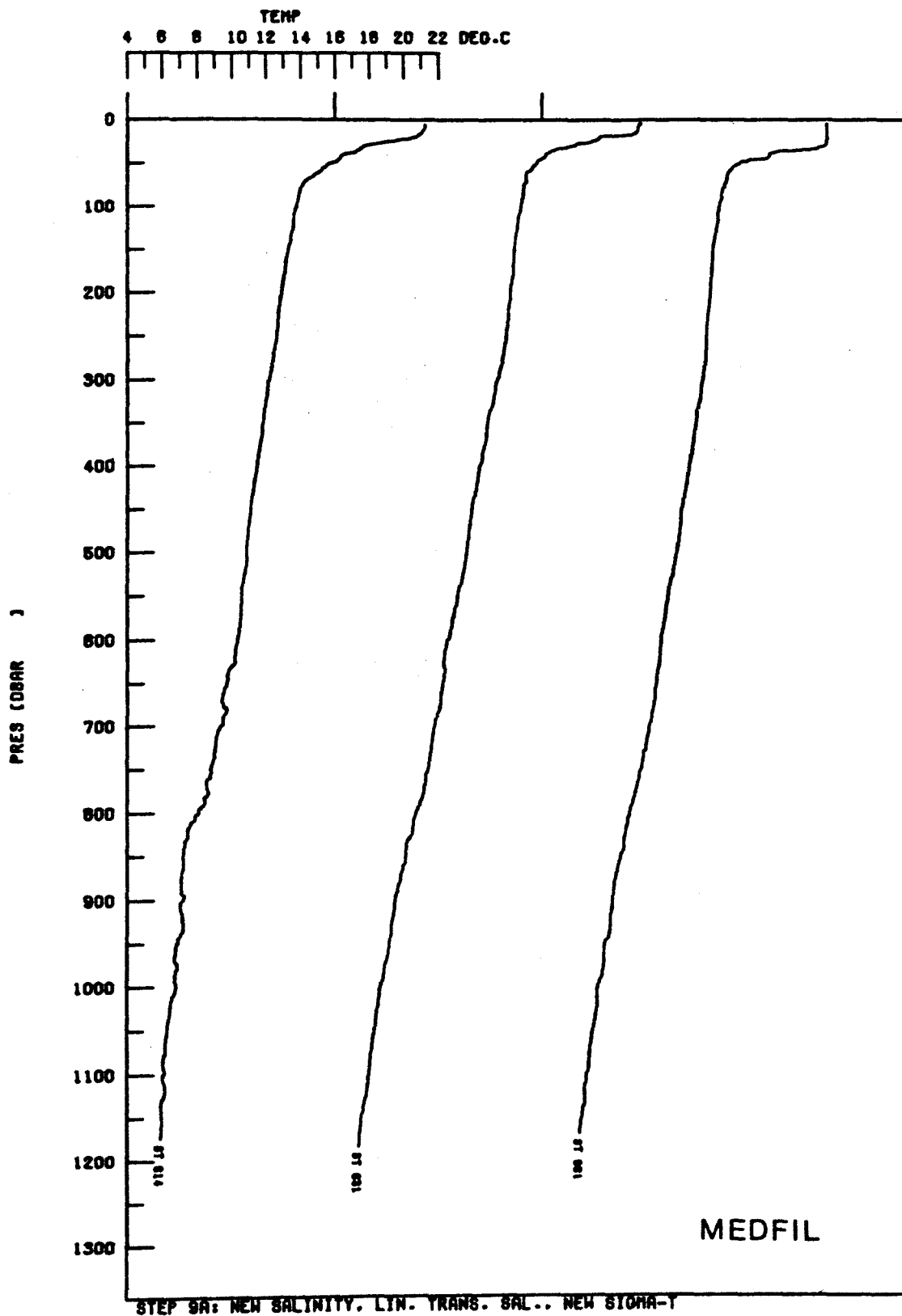
Fig. 12a-d: Temperature data after passing stage 2 with different editing techniques.

MEDDIF

STEP 9B: NEW SALINITY. LIN. TRANS. SAL.. NEW SIGMA-T

Fig. 12b

STADEV

STEP 8C: NEW SALINITY, LIN. TRANS. SAL., NEW SIGMA-T
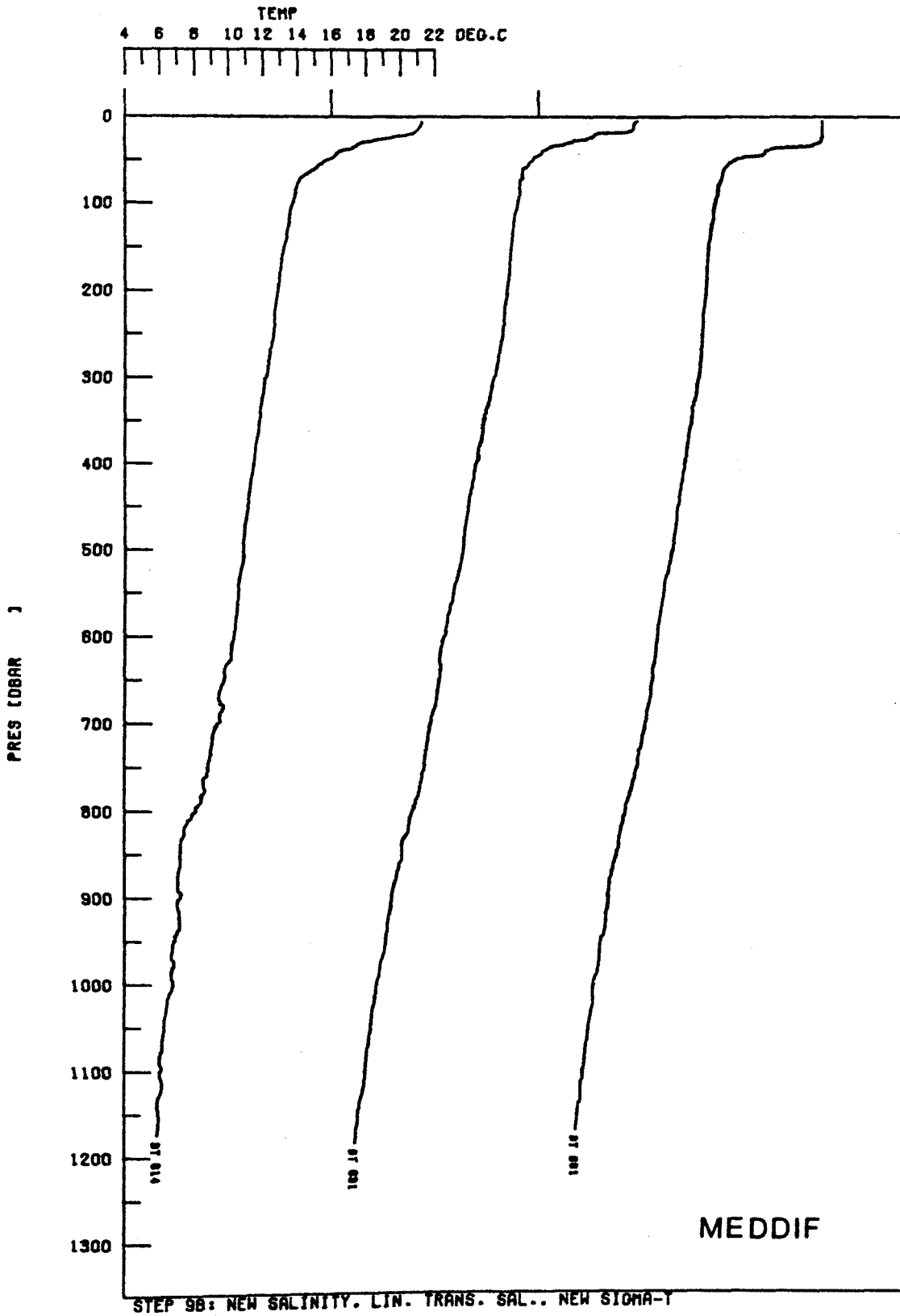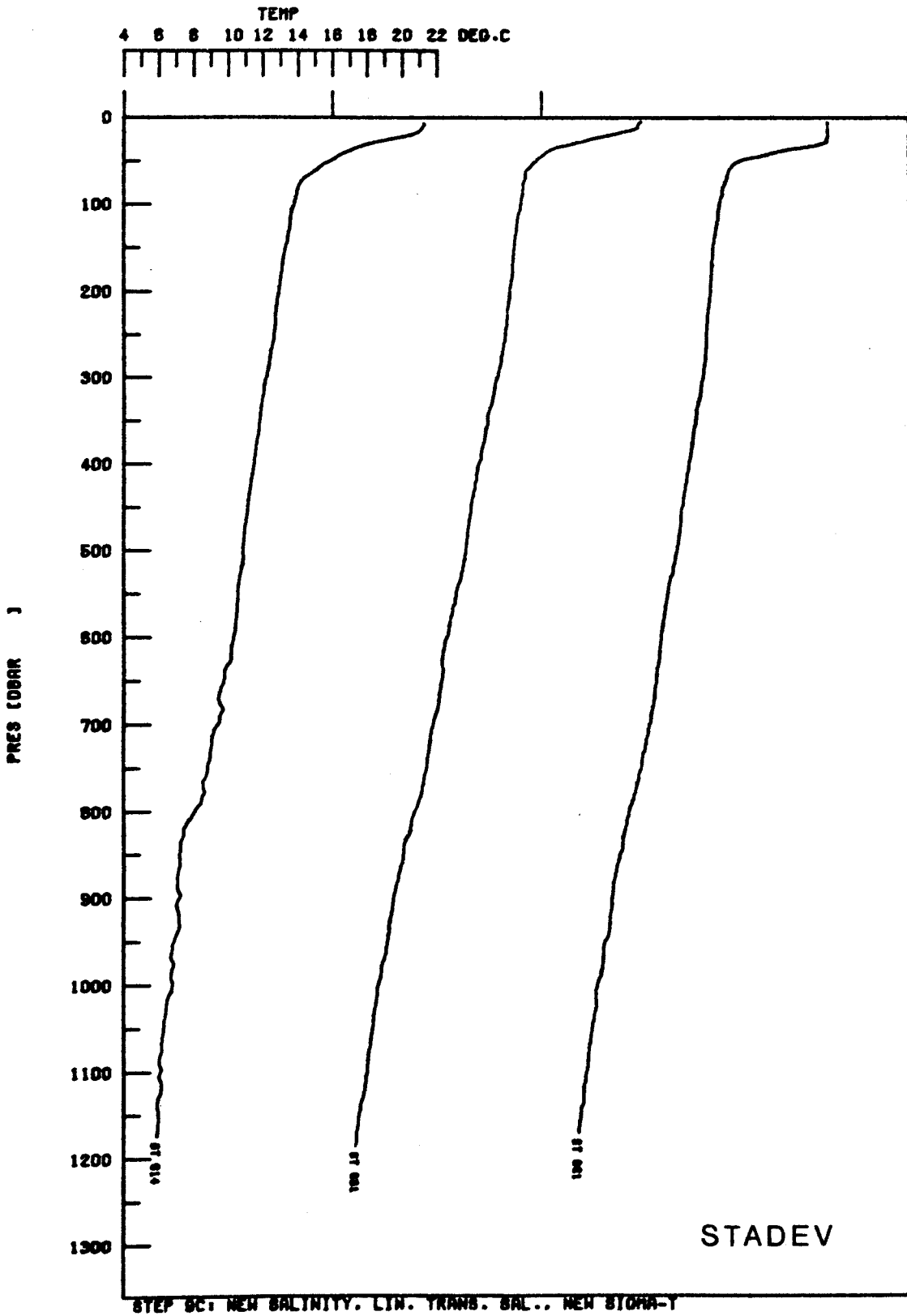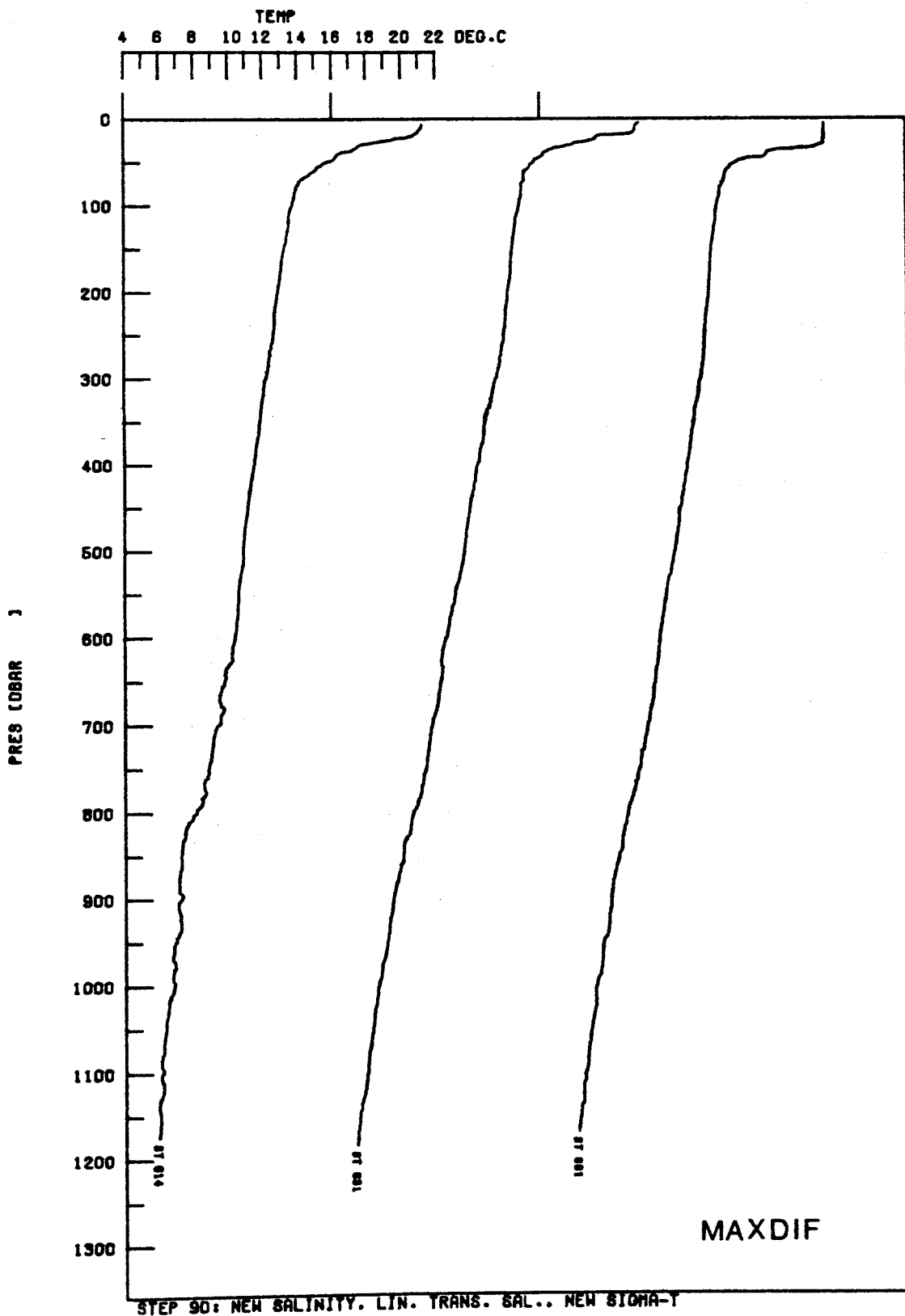
Fig. 12c

Fig. 12d
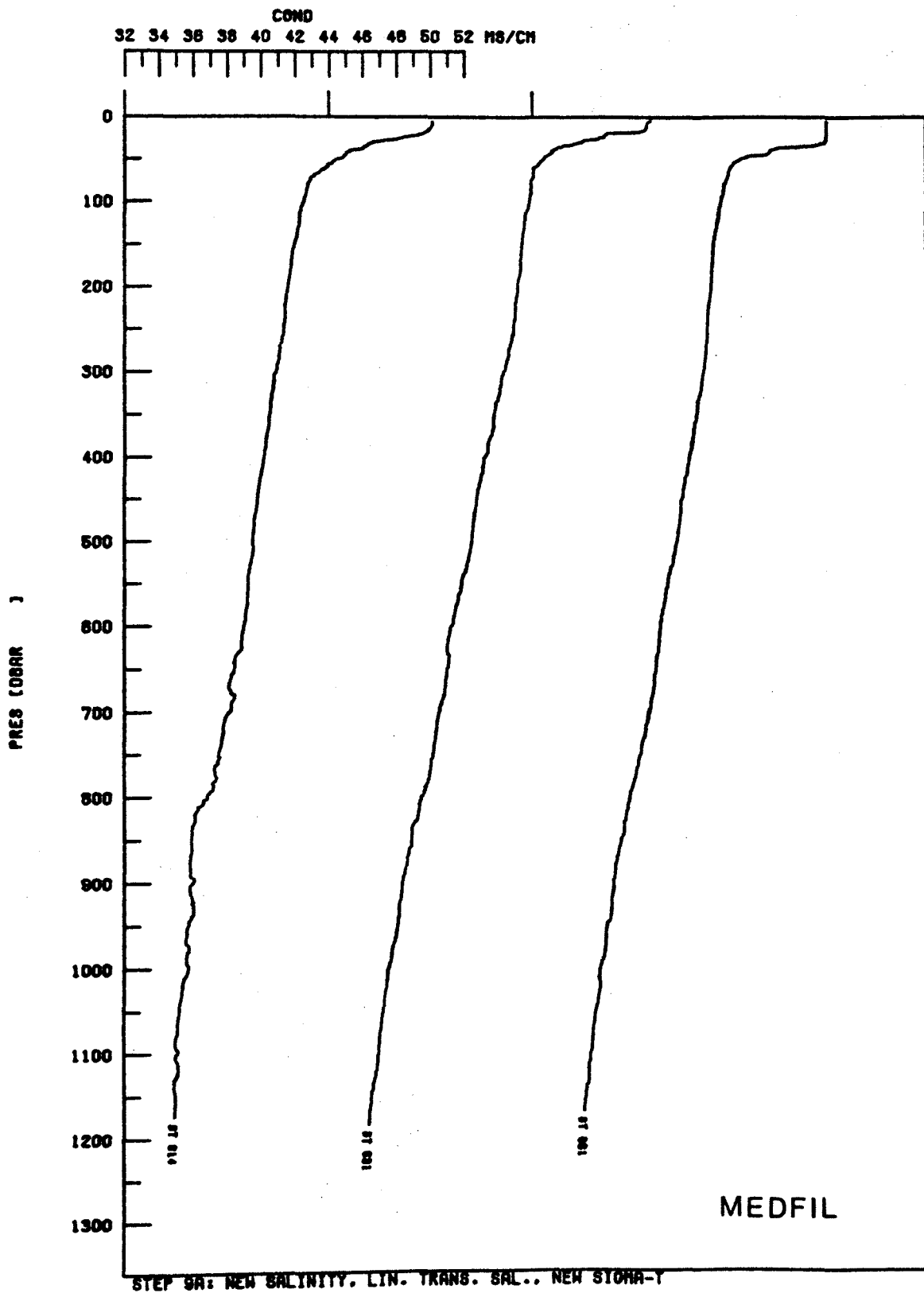
Fig. 13a-d: Conductivity data after passing stage 2 with different editing techniques.
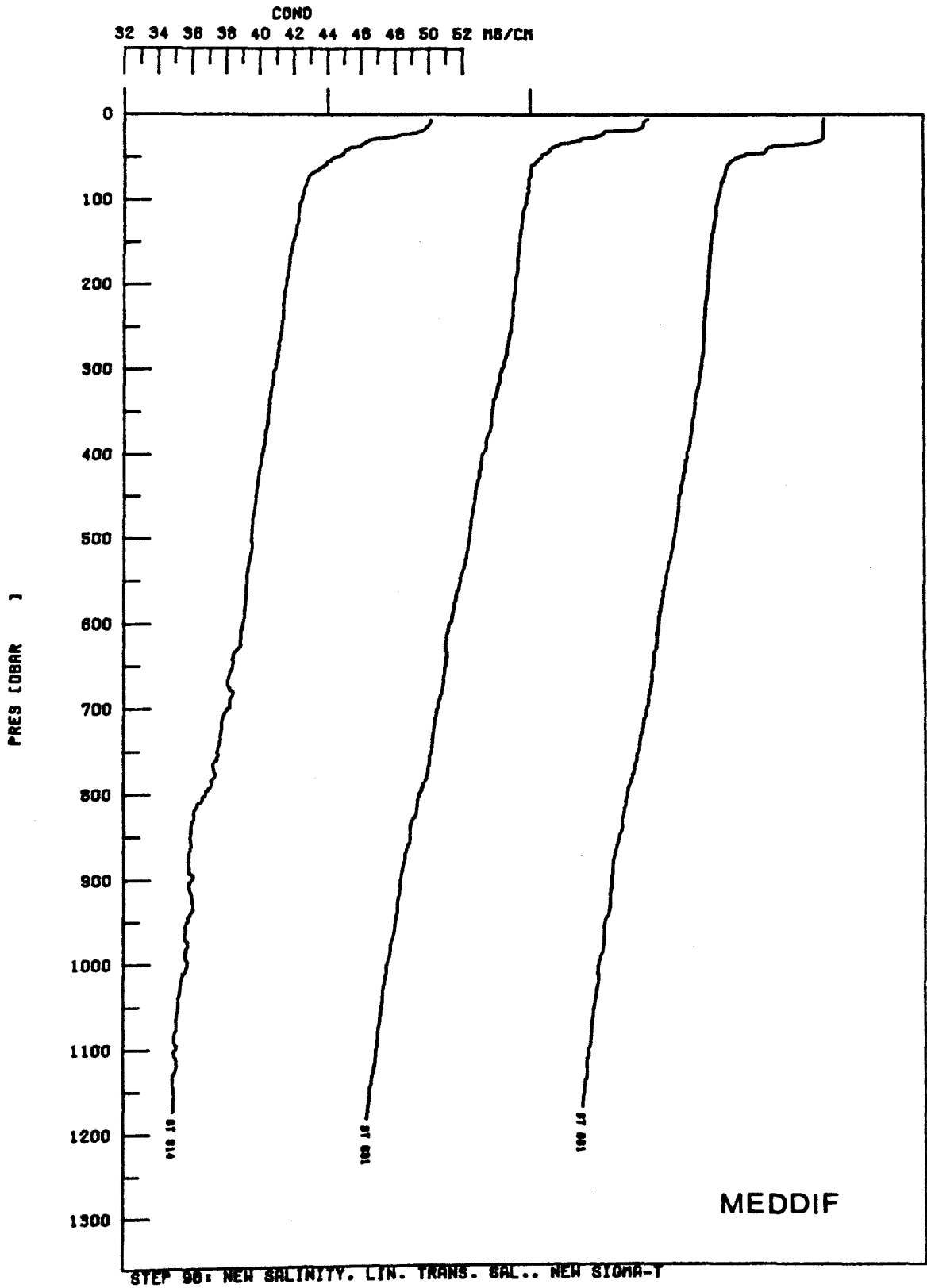
Fig. 13b
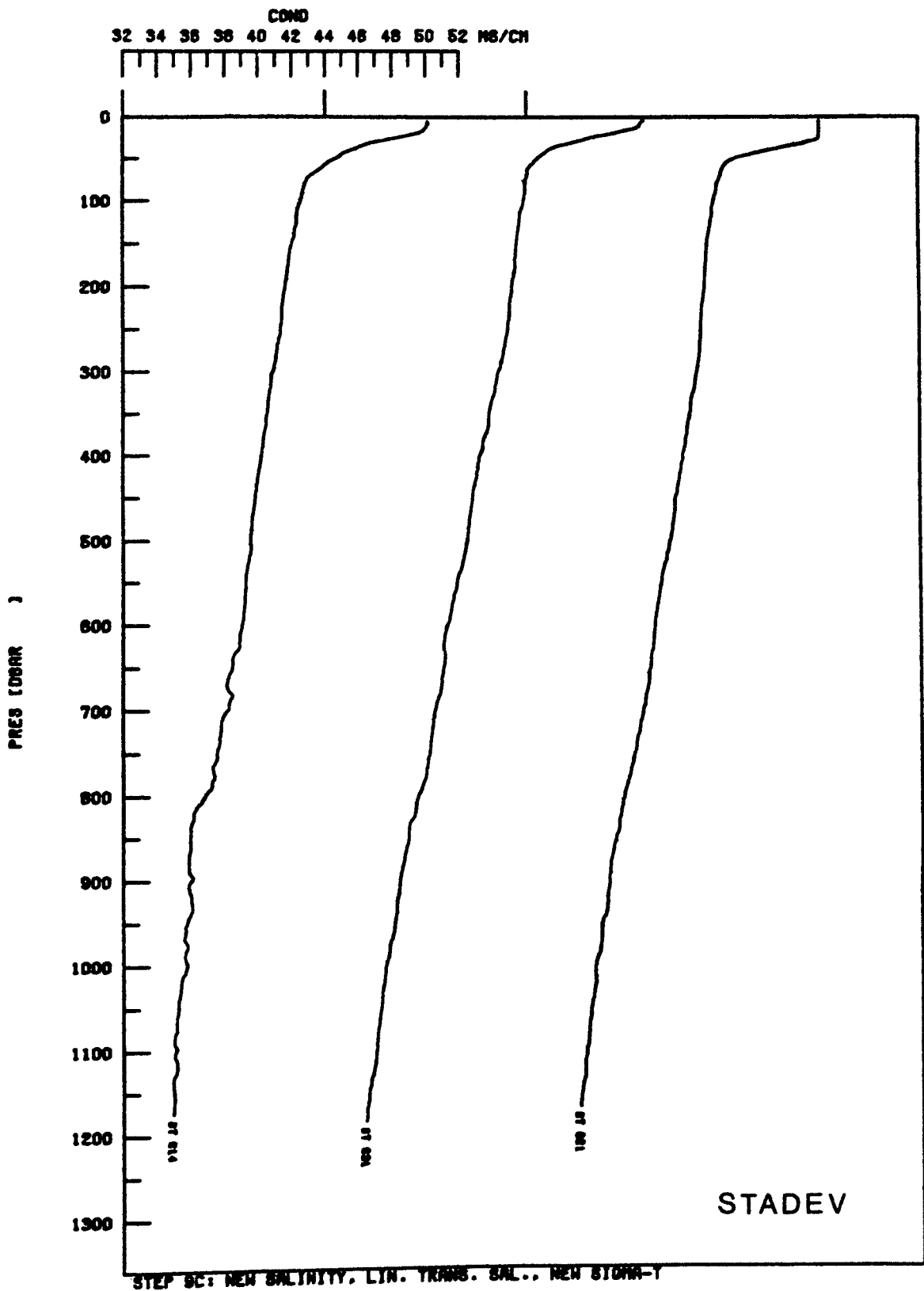
Fig. 13c
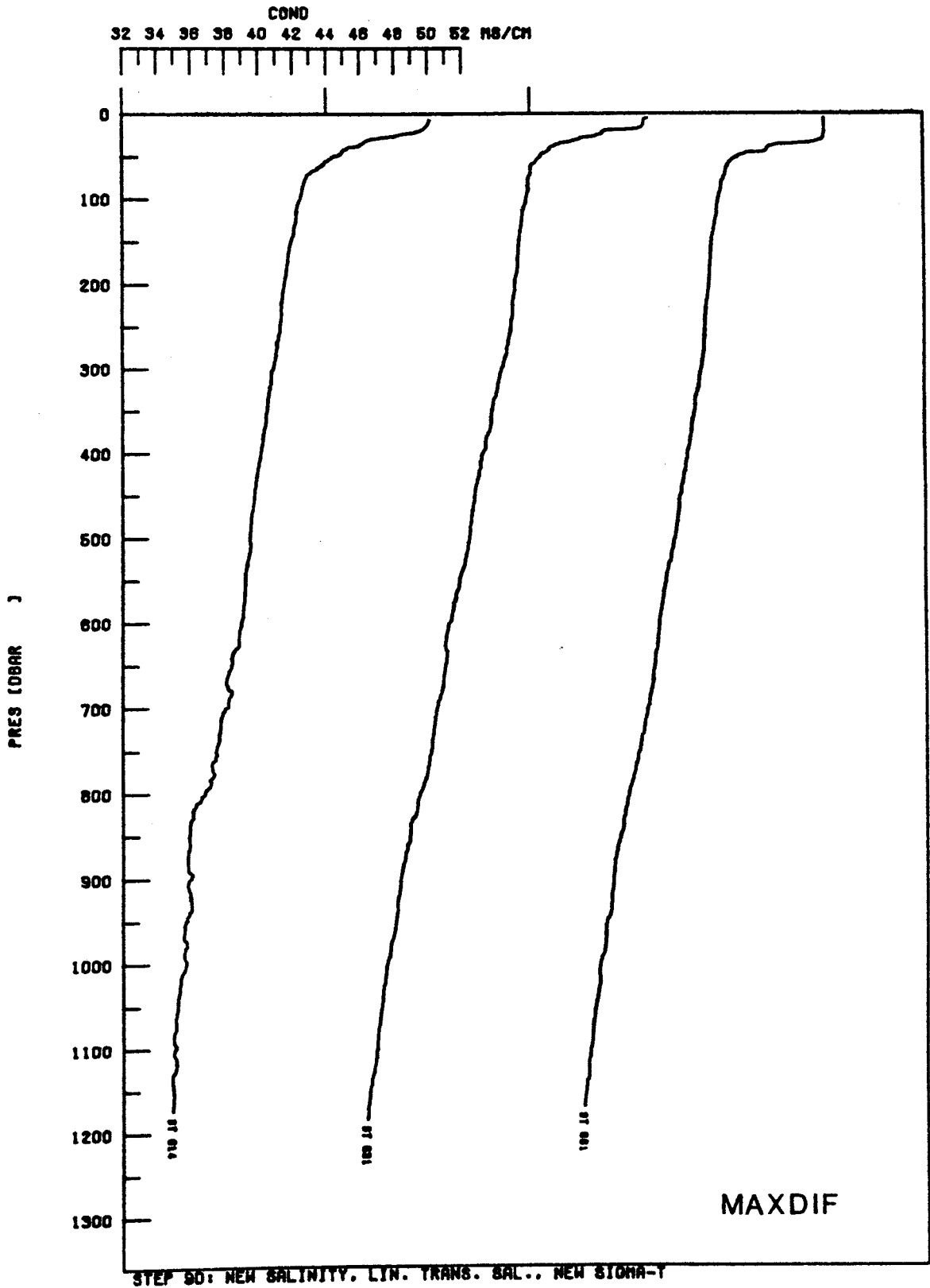
Fig. 13d
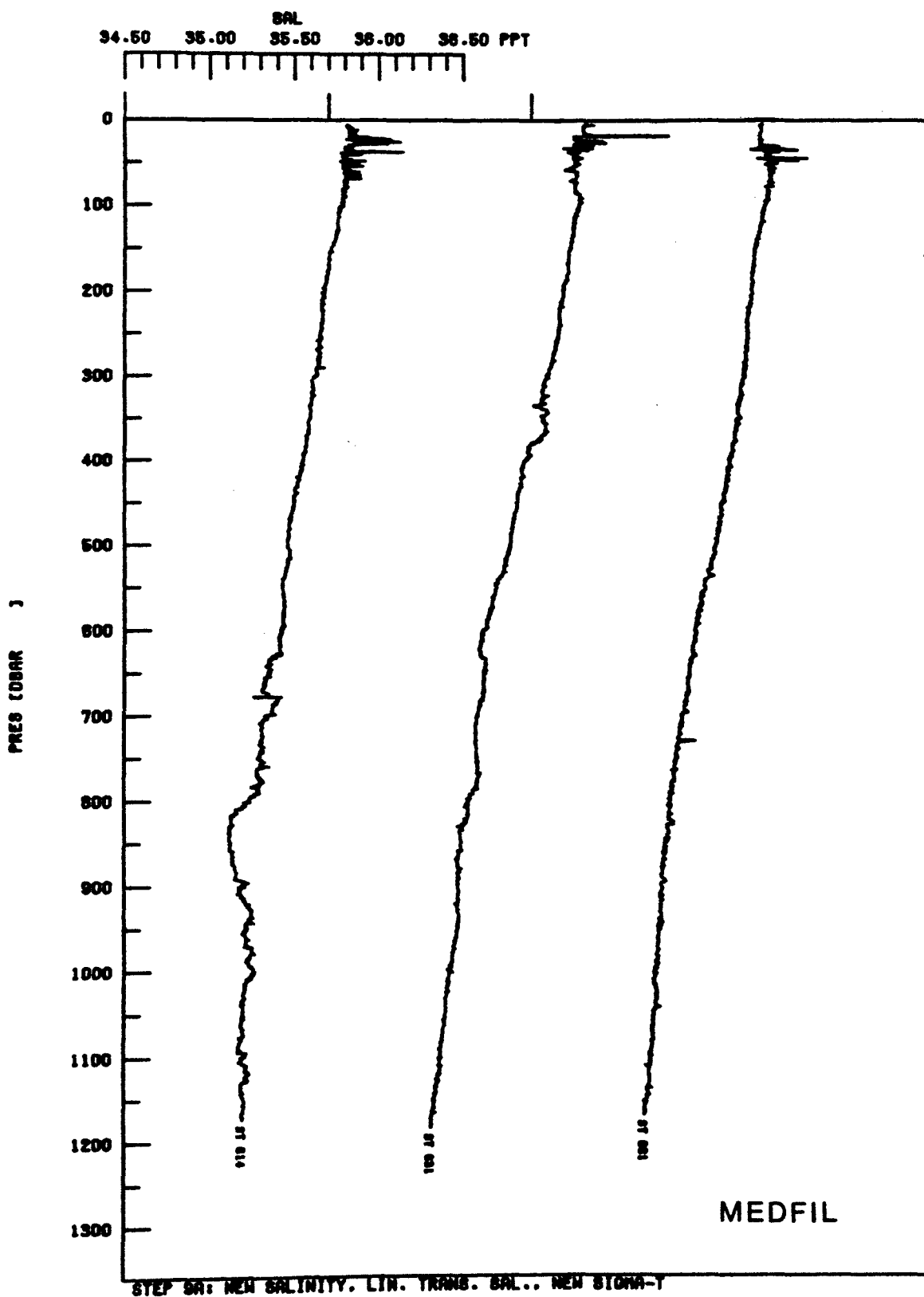
Fig. 14a-d: Salinity data after passing stage 2 with different editing techniques.
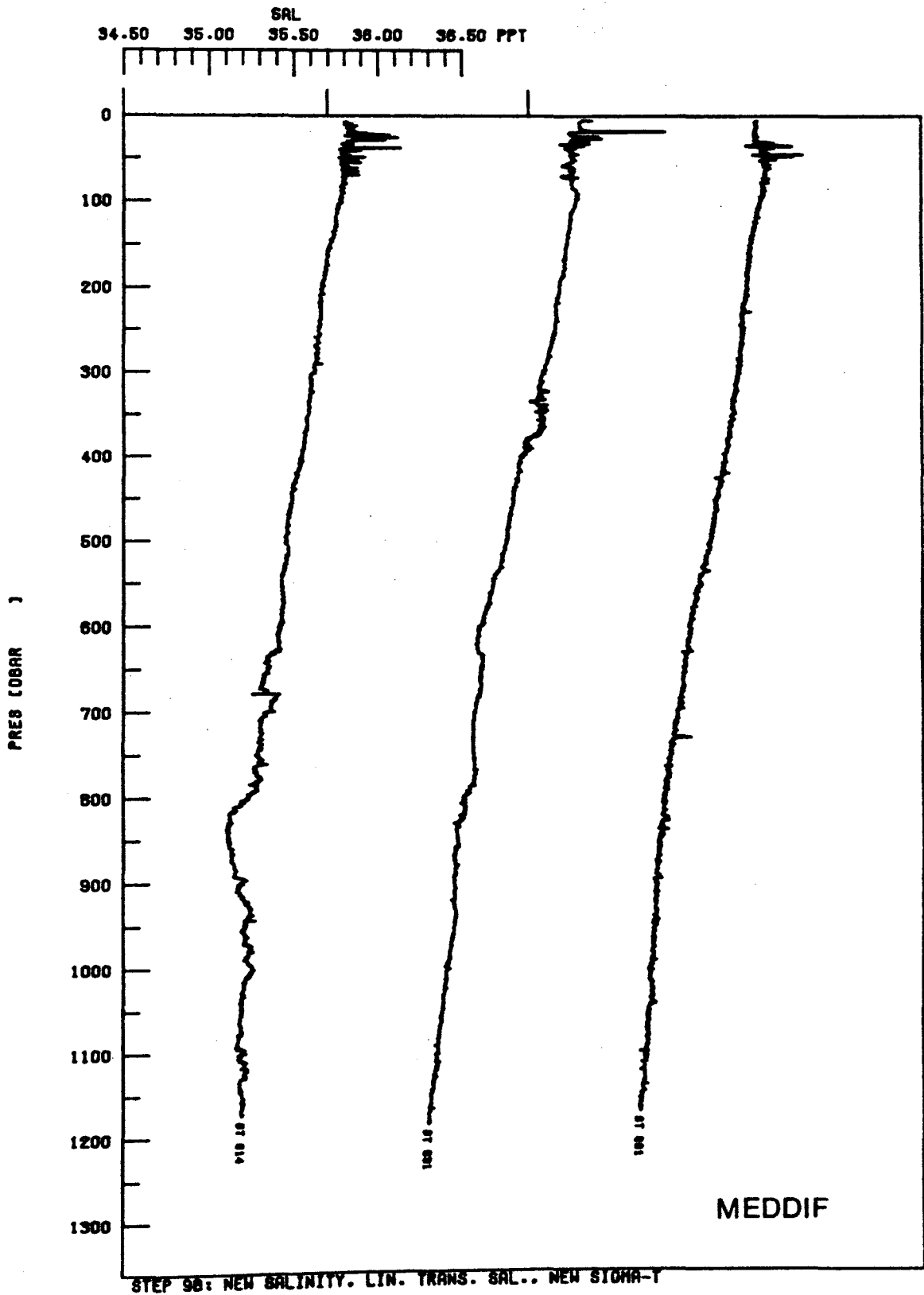
Fig. 14b
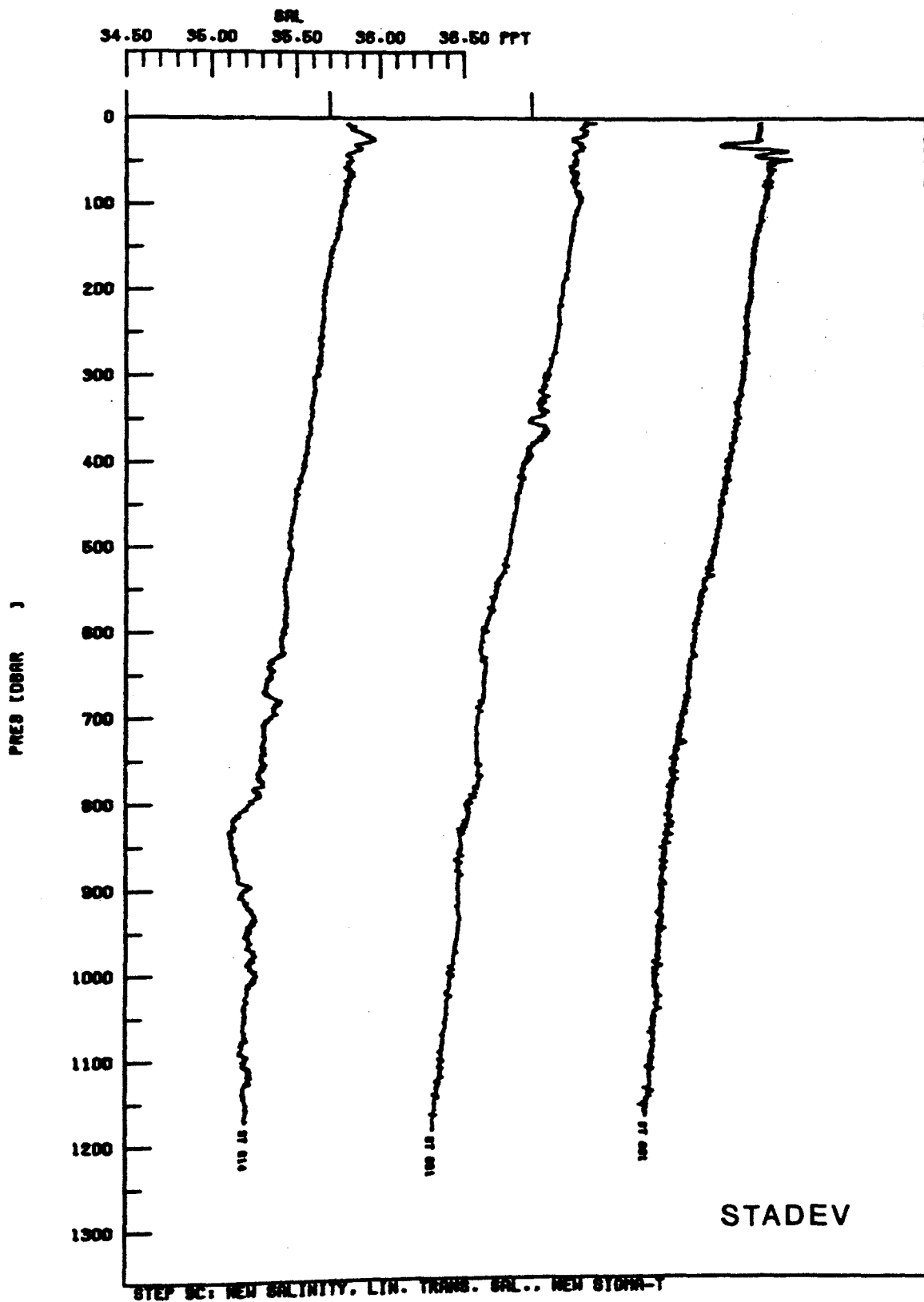
Fig. 14c
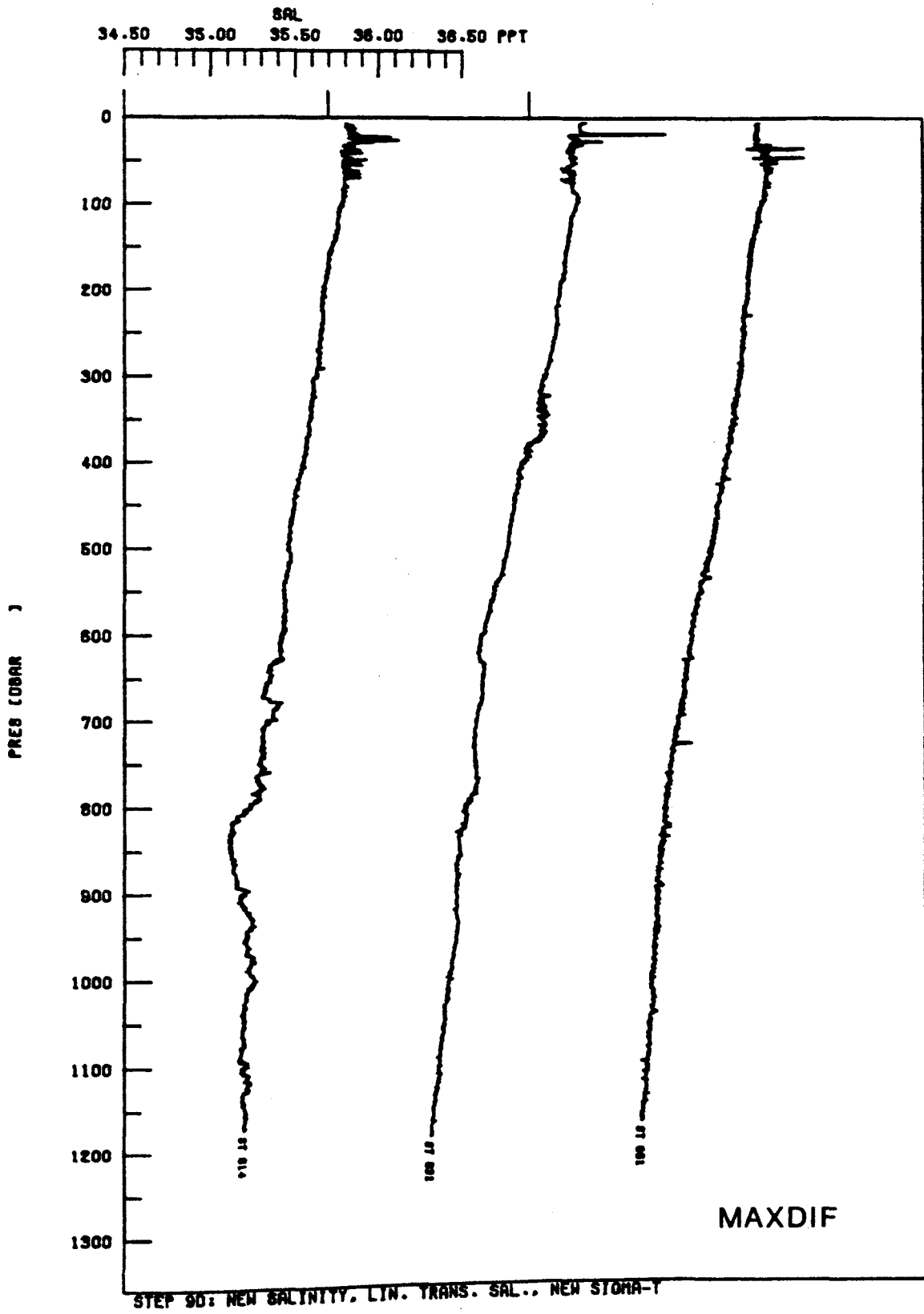
Fig. 14d
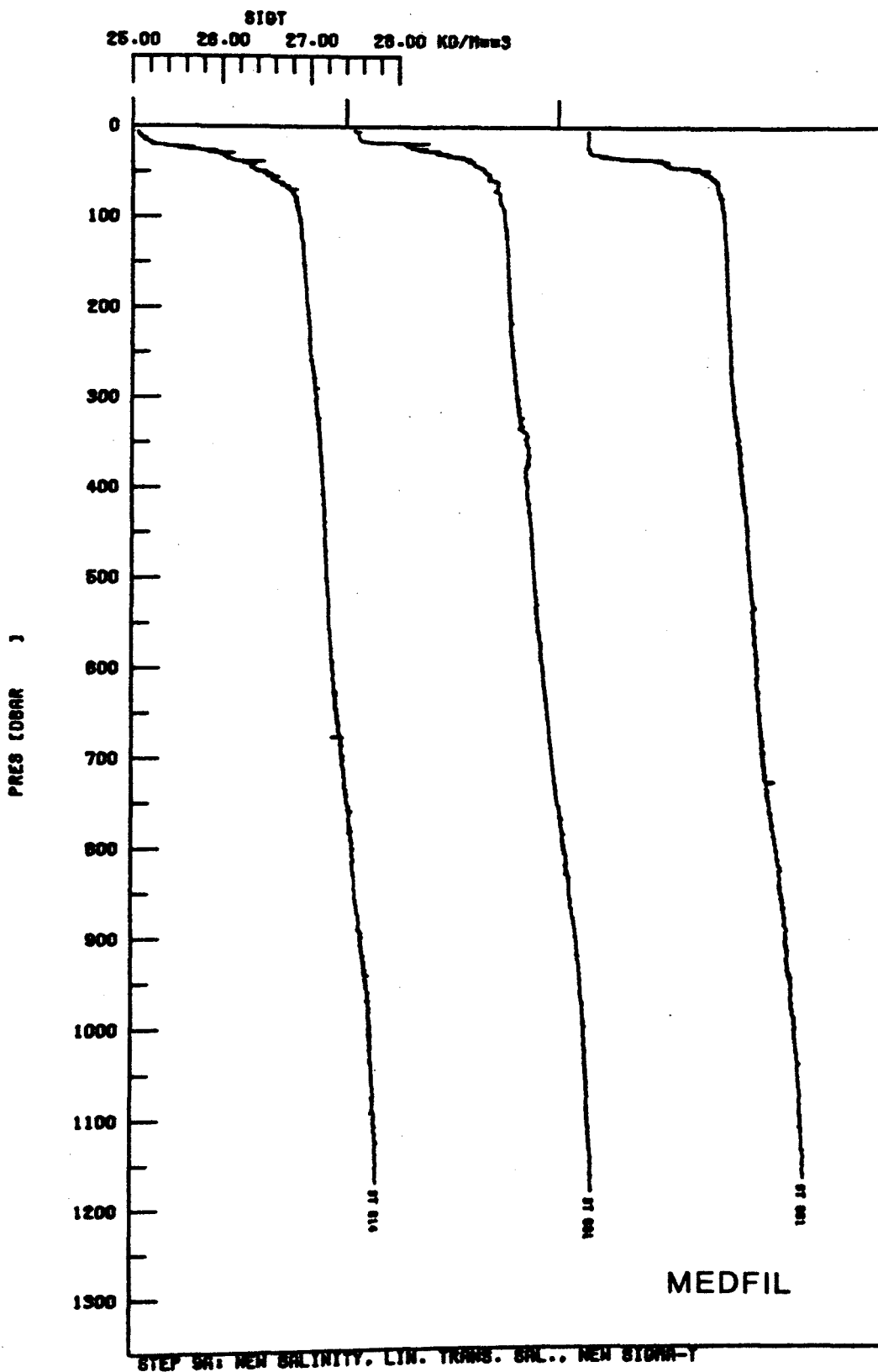
Fig. 15a-d: Density data after passing stage 2 with different editing techniques.

MEDDIF

STEP 98: NEW SALINITY, LIN. TRANS. SAL., NEW SIGMA-T
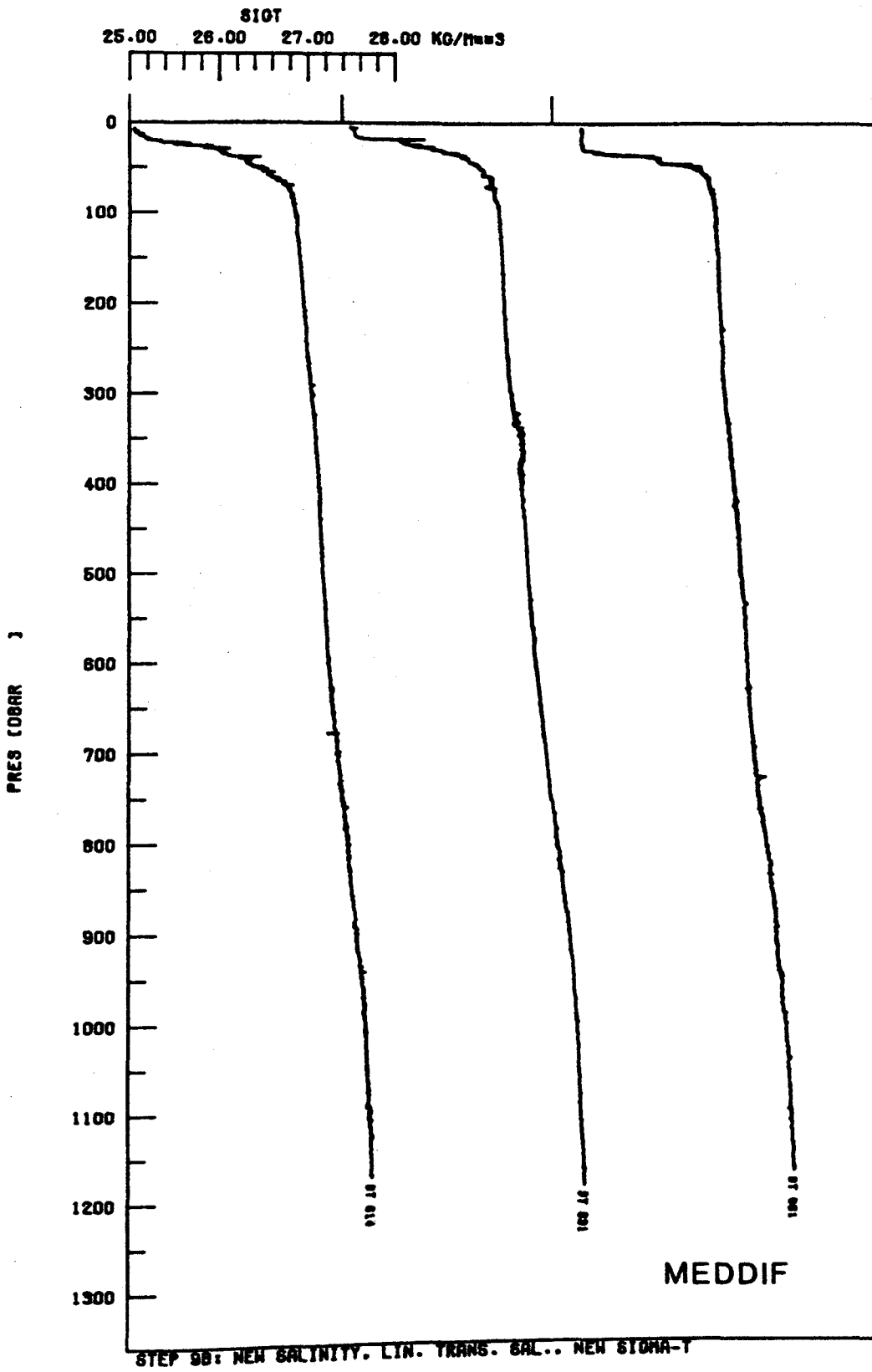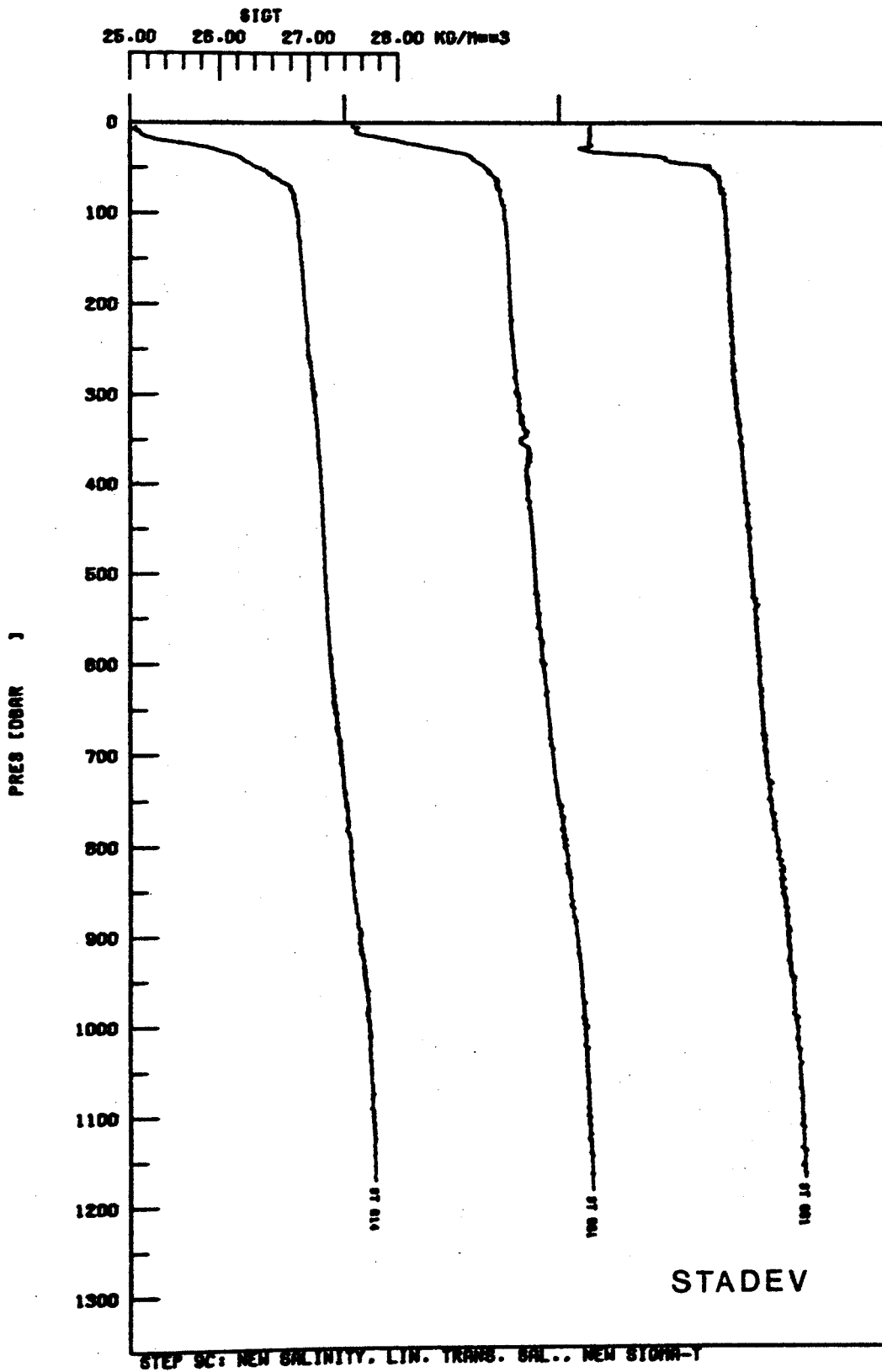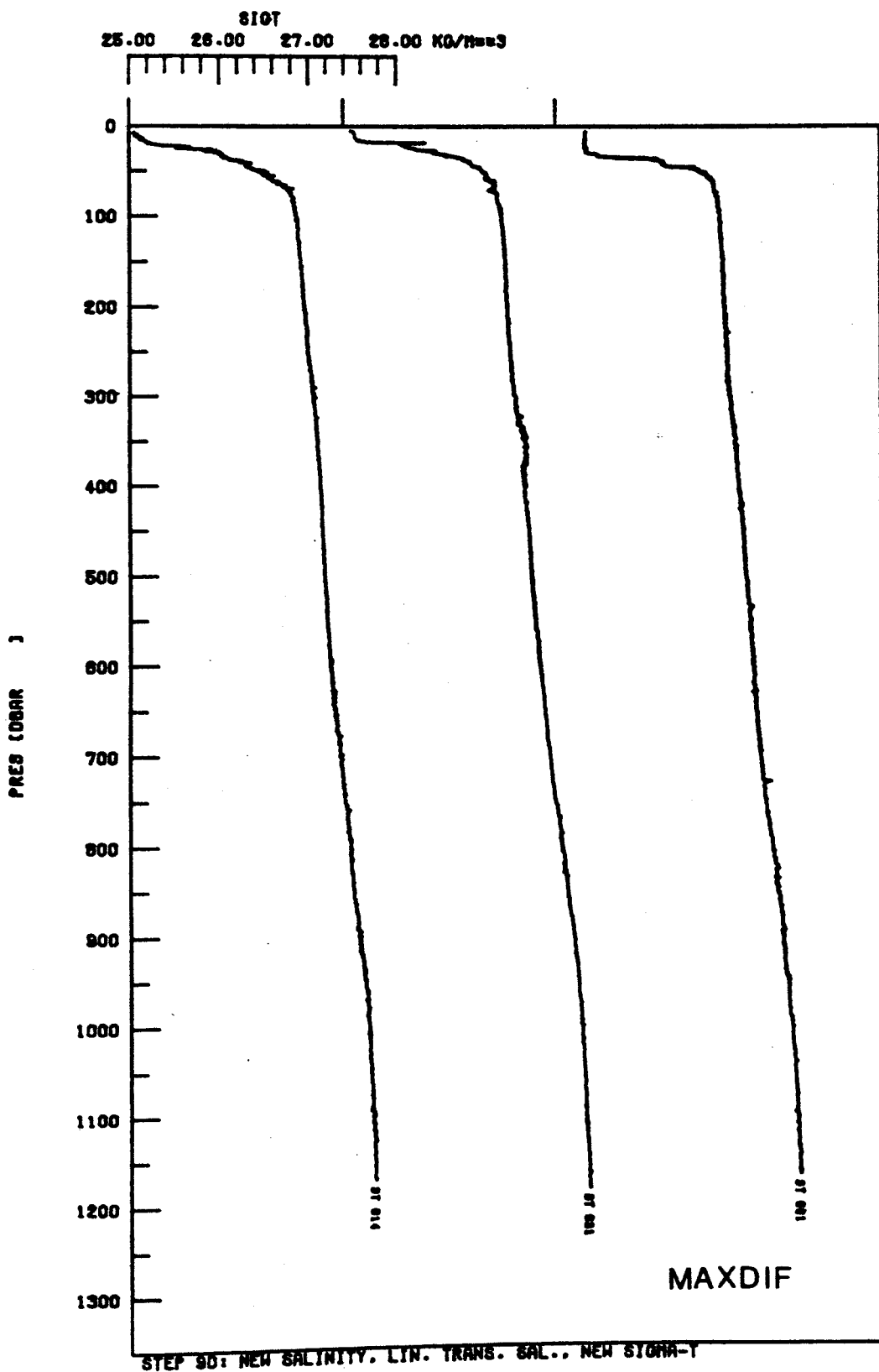
Fig. 15b

Fig. 15c

Fig. 15d
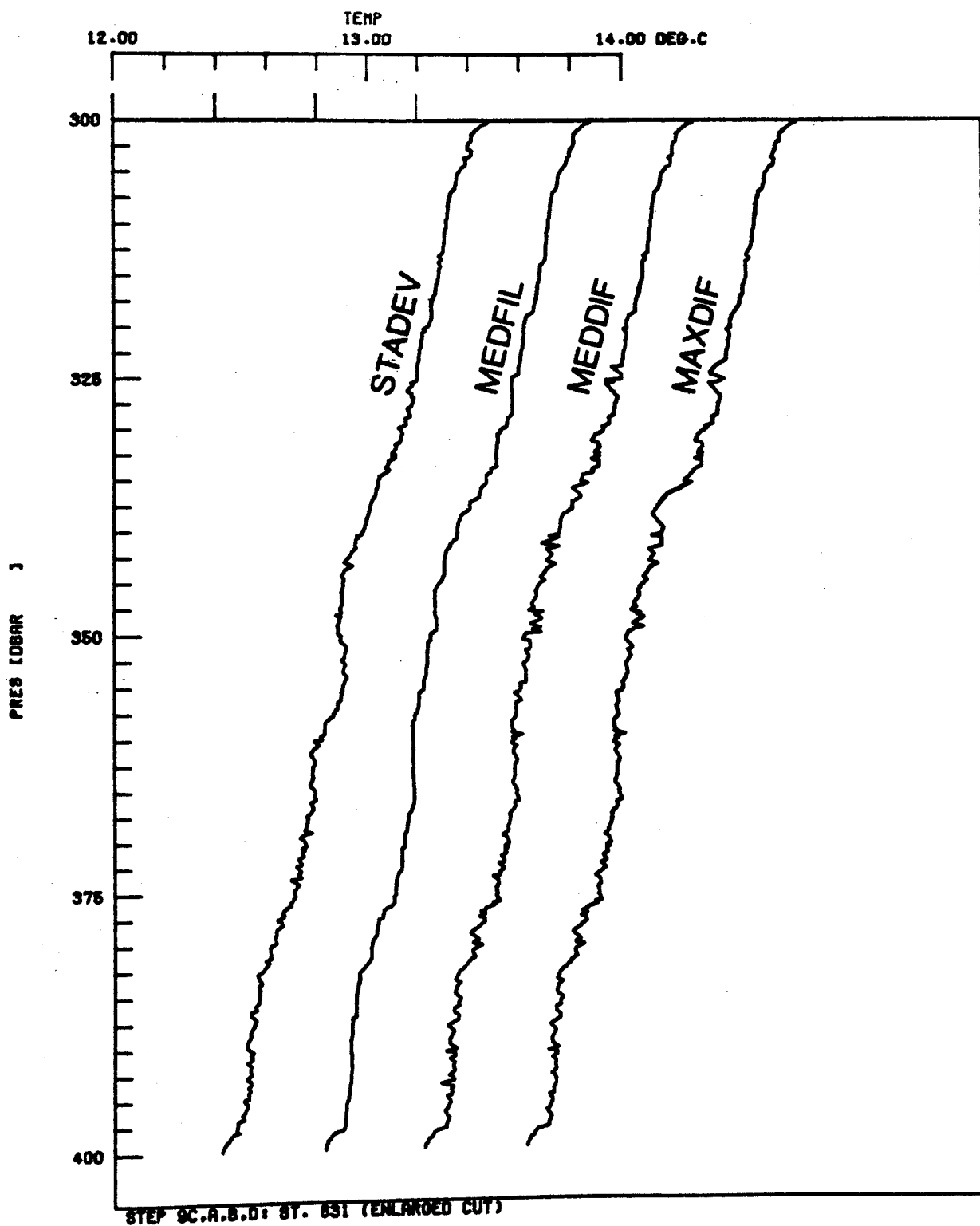
Fig. 16: Enlarged cut of temperature data of the spiky part of St. 631 from fig. 8a, comparing all four editing techniques after stage 2.
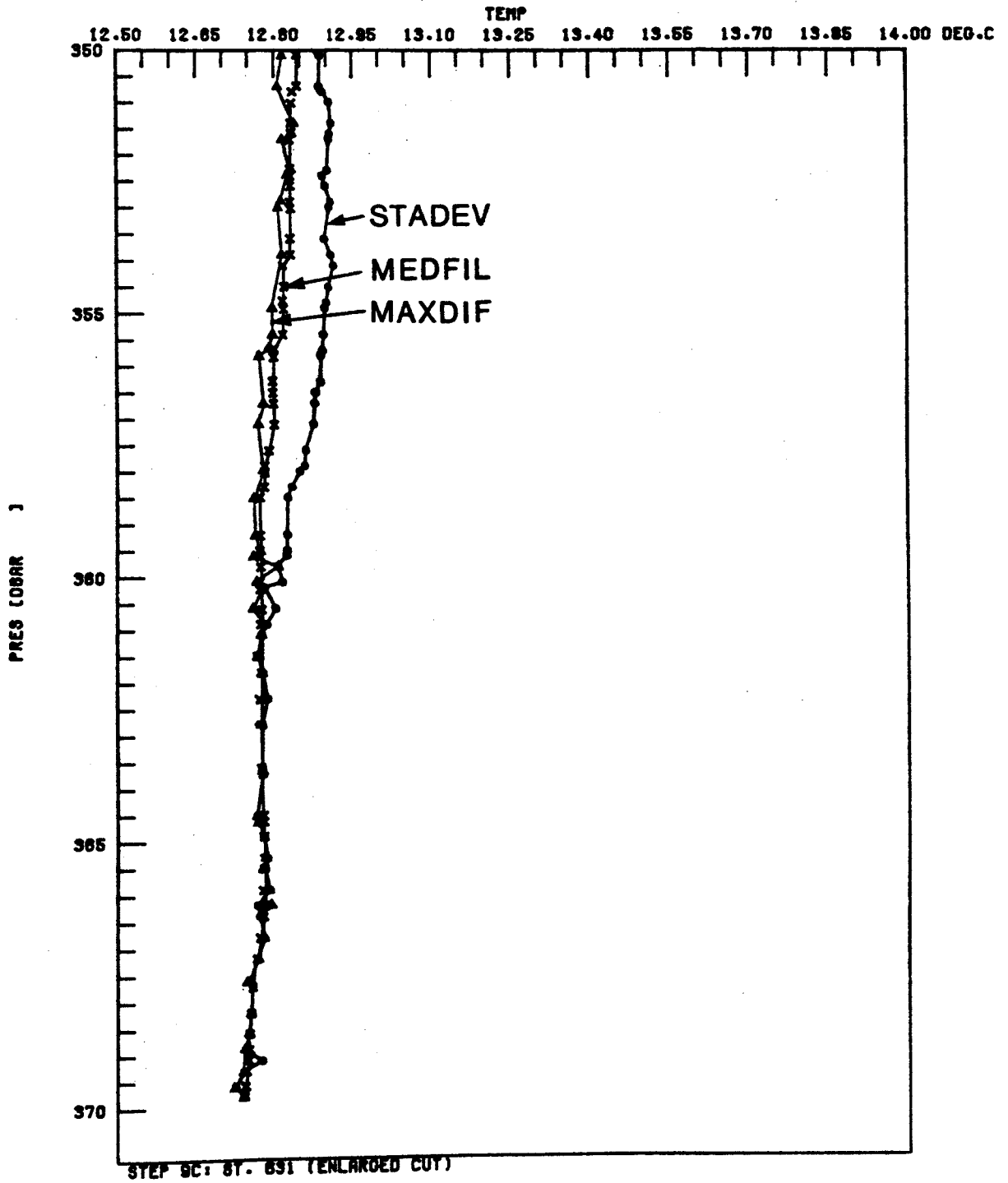
Fig. 17:   Same as fig. 16, but with data point resolution.
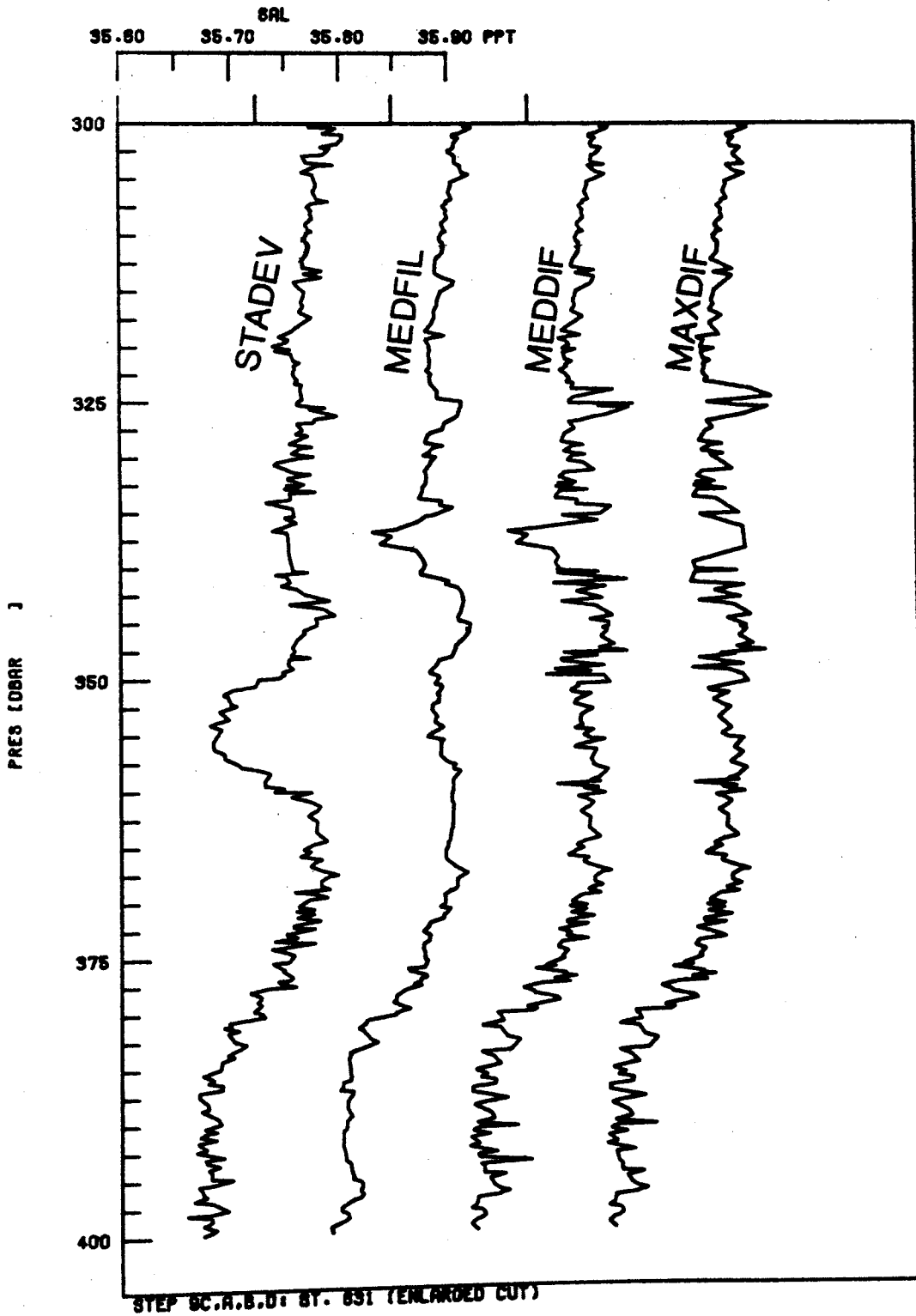
Fig. 18a: Enlarged cut of salinity data of the spiky part of St. 631 from fig. 8a, after passing stage 2.
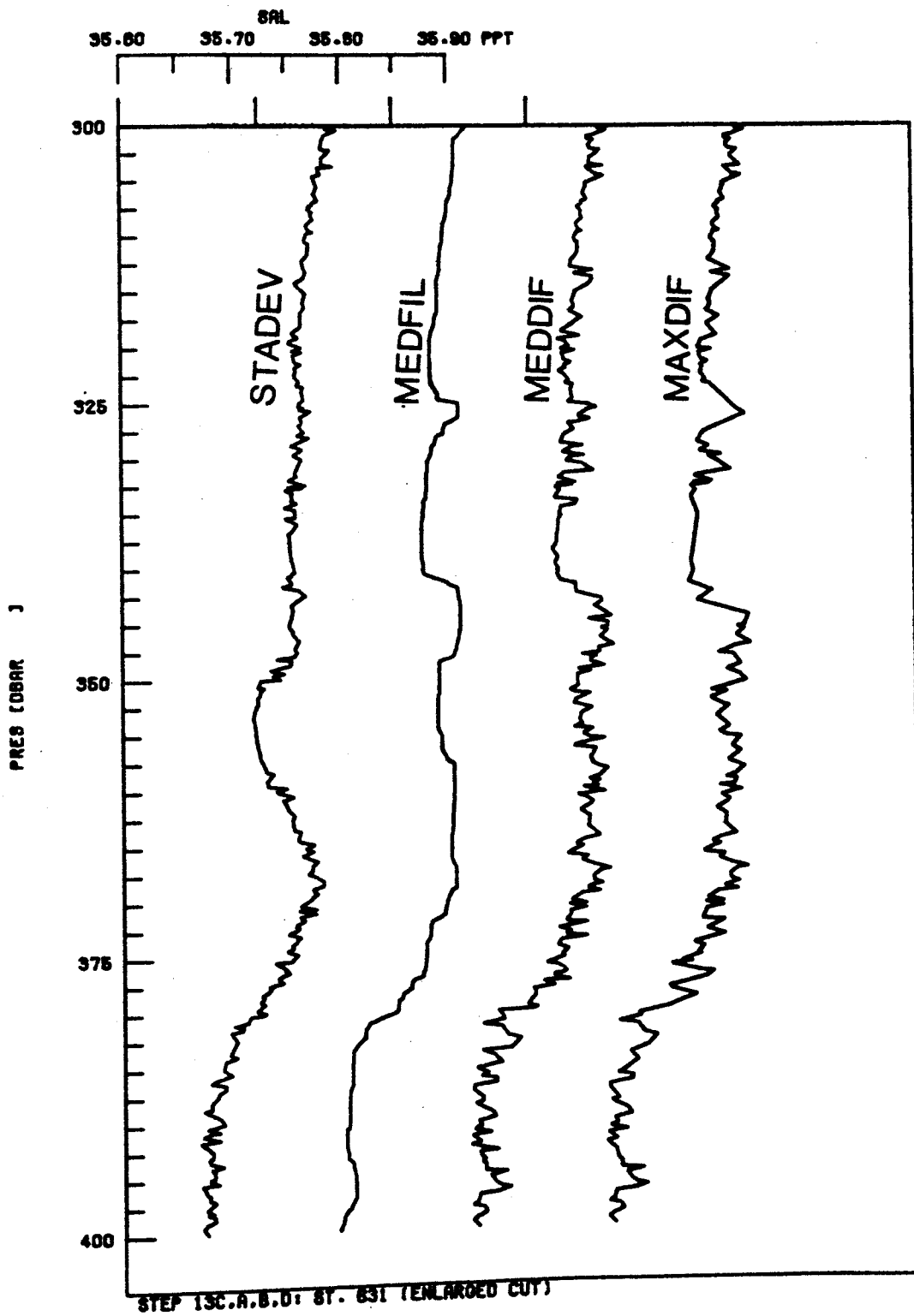
Fig. 18b: Same as fig. 18a, but after passing stage 3.
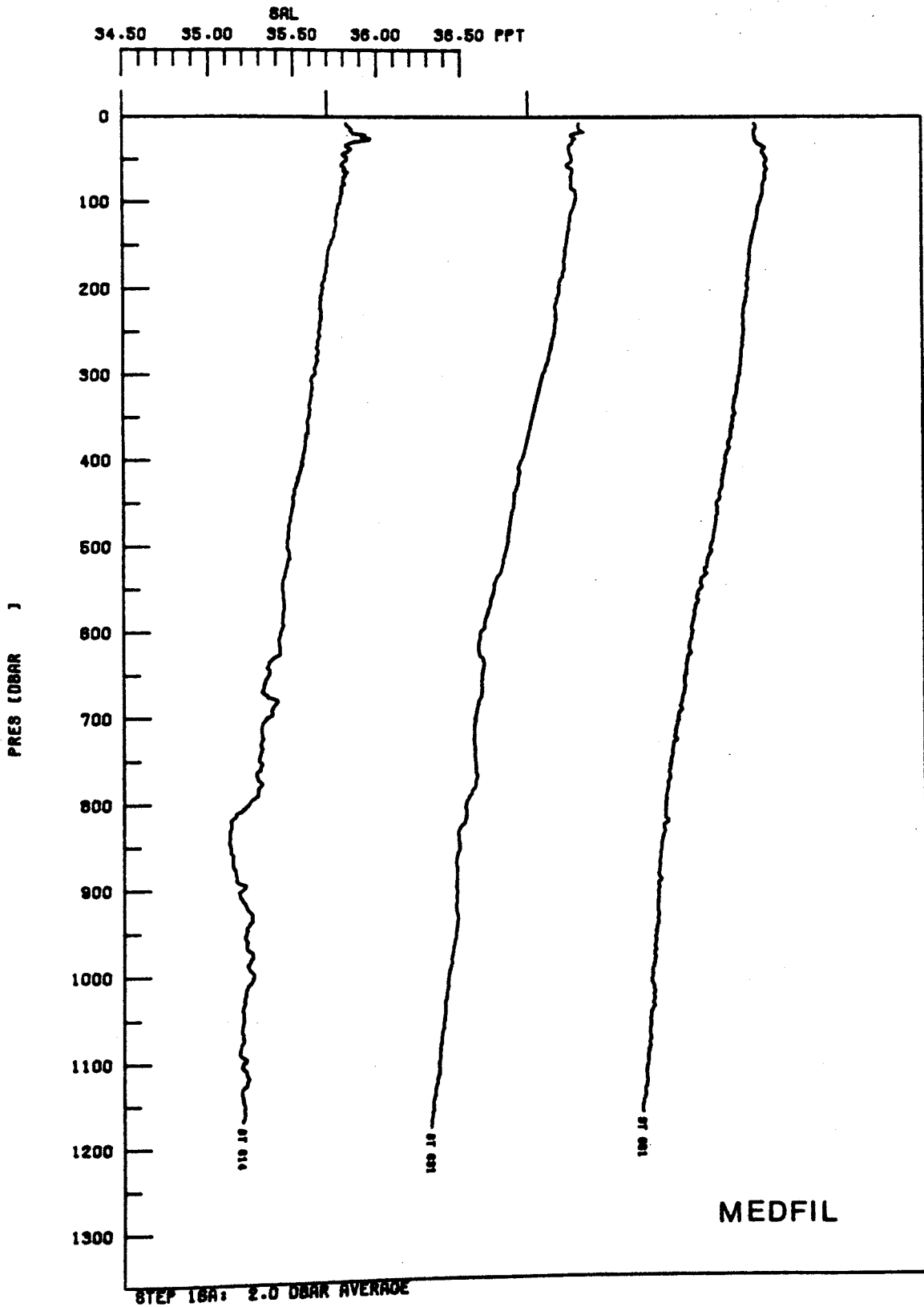
Fig. 19a-d: Final salinity data for the different editing
techniques.
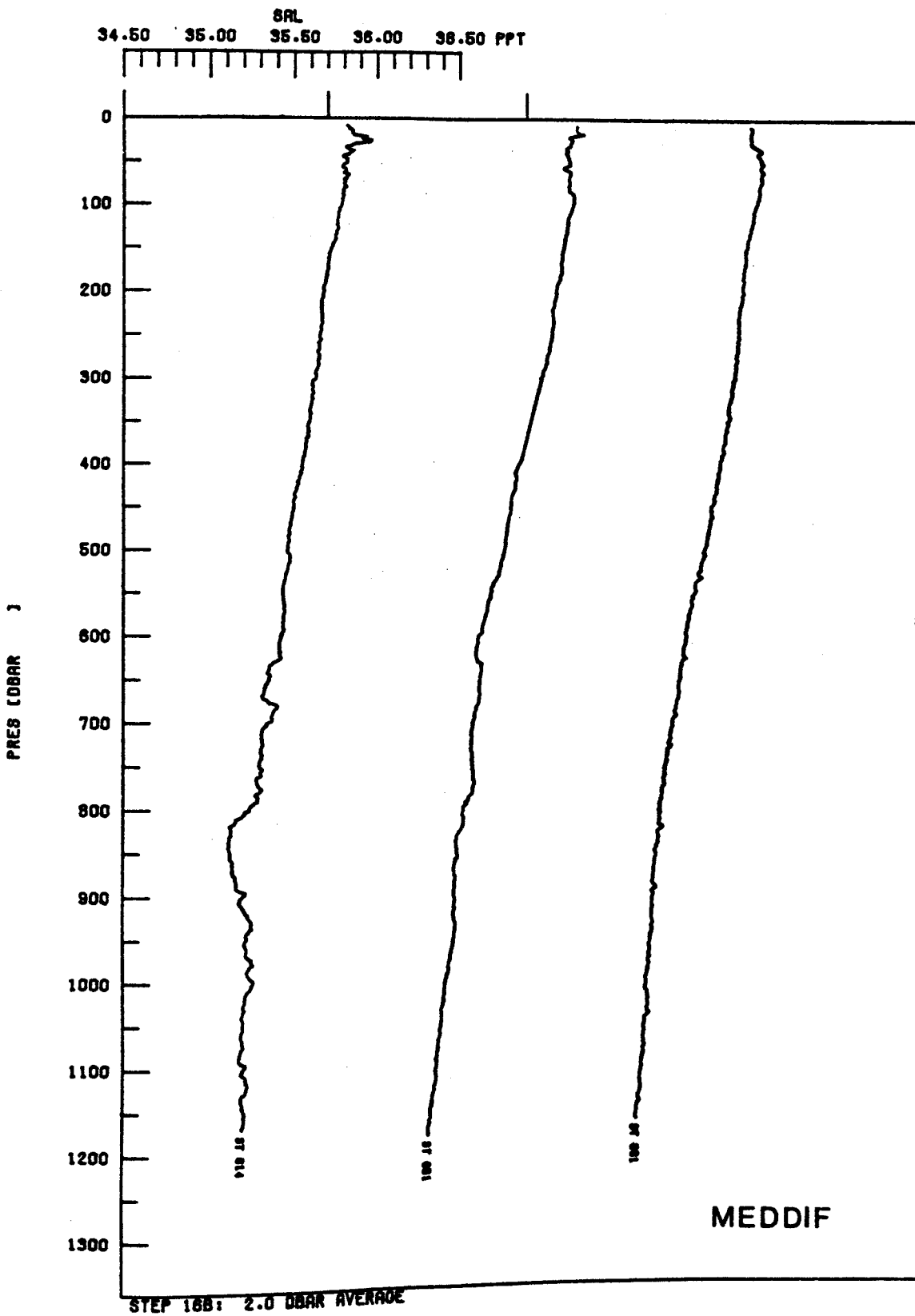
Fig. 19b
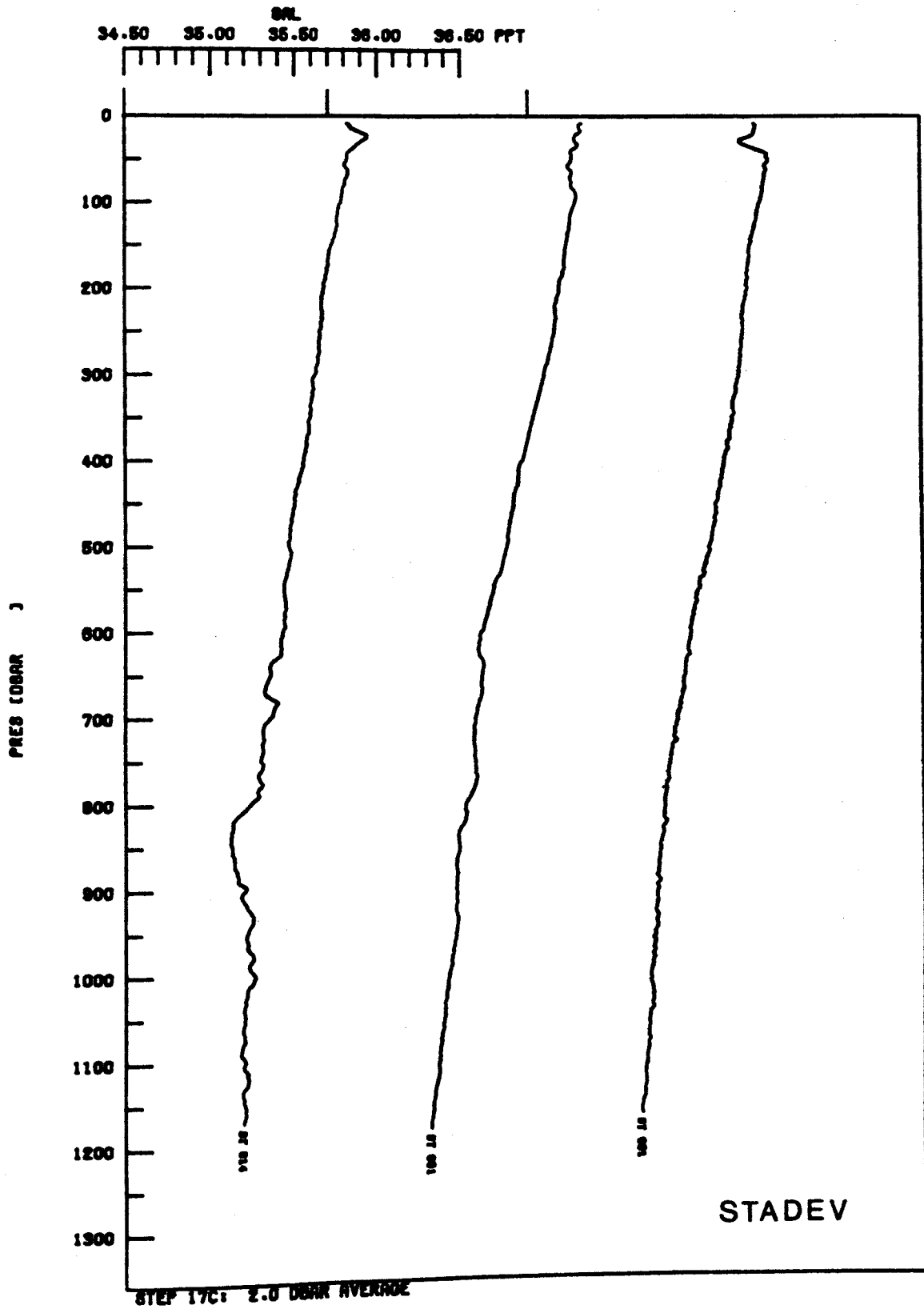
Fig. 19c
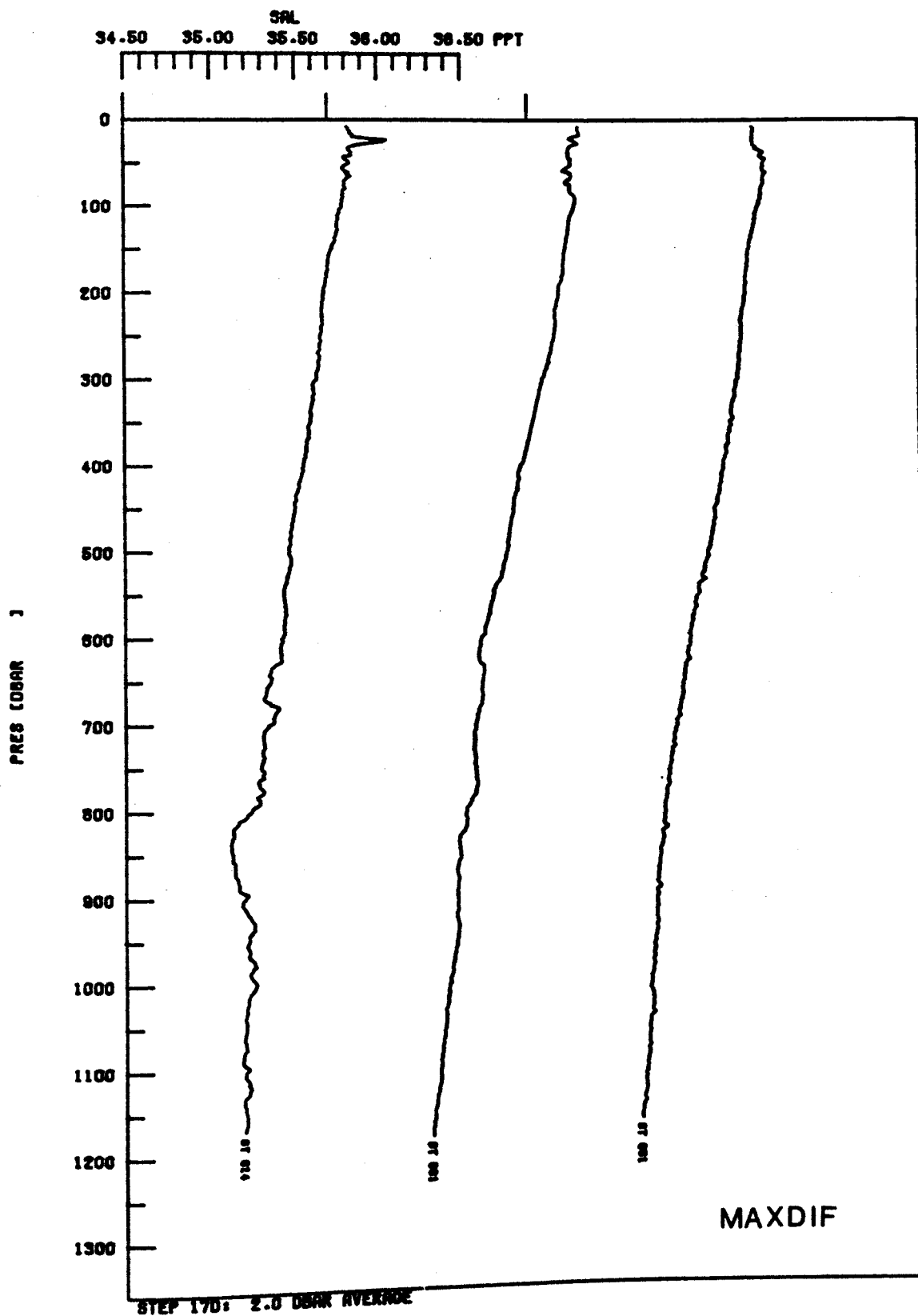
Fig. 19d
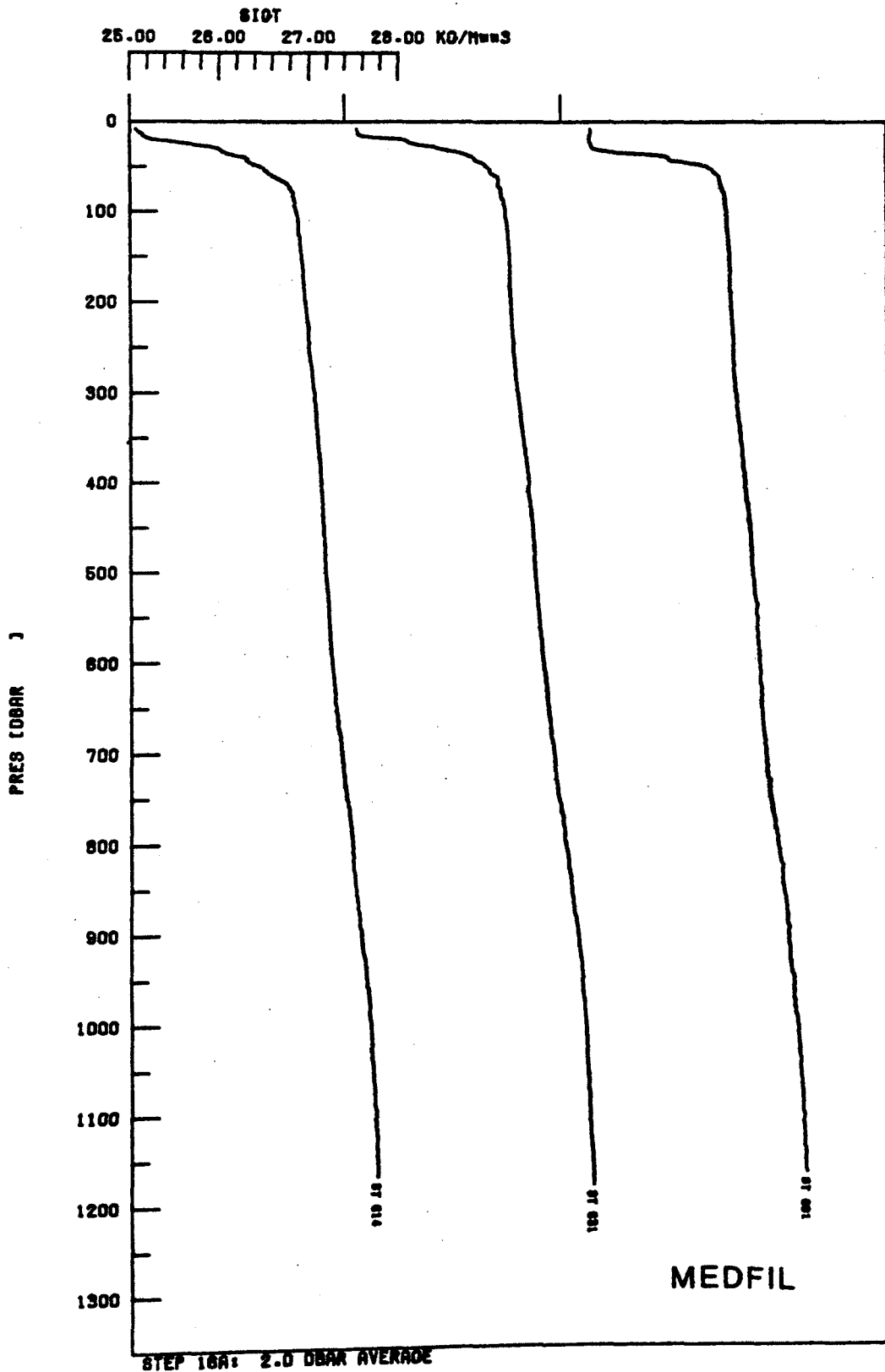
Fig. 20a-d: Final density data for the different editing techniques.
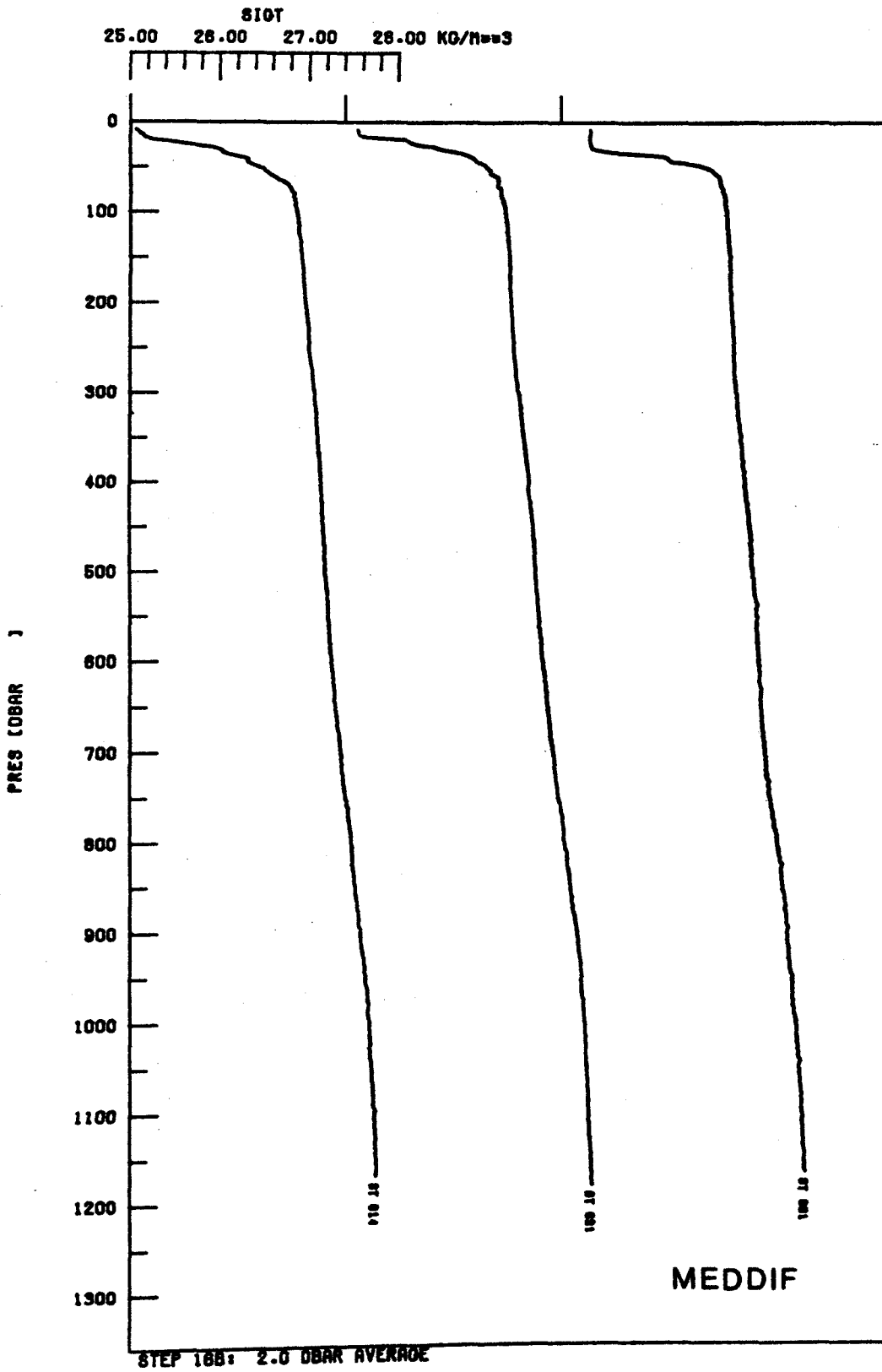
Fig. 20b
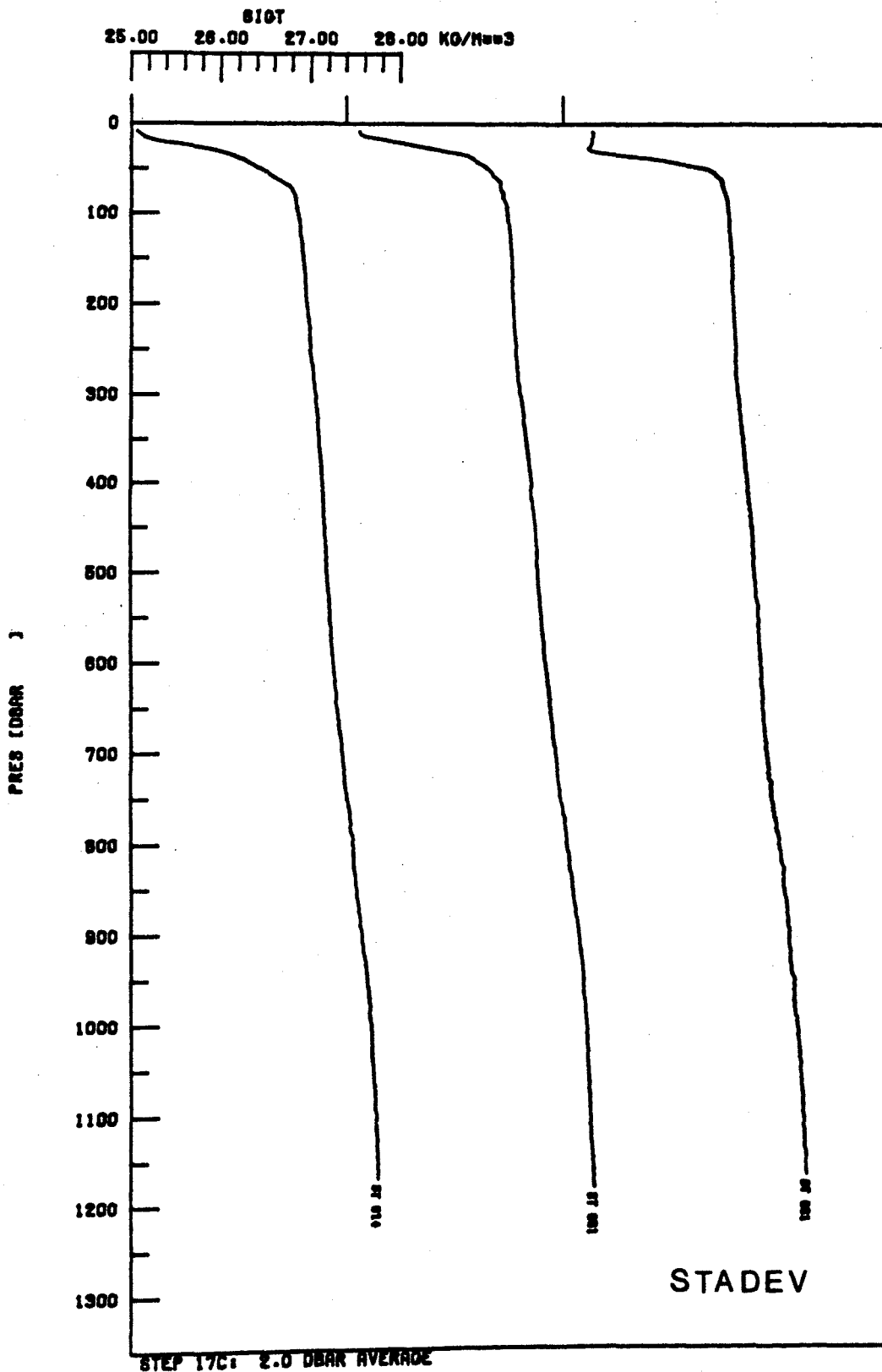
Fig. 20c
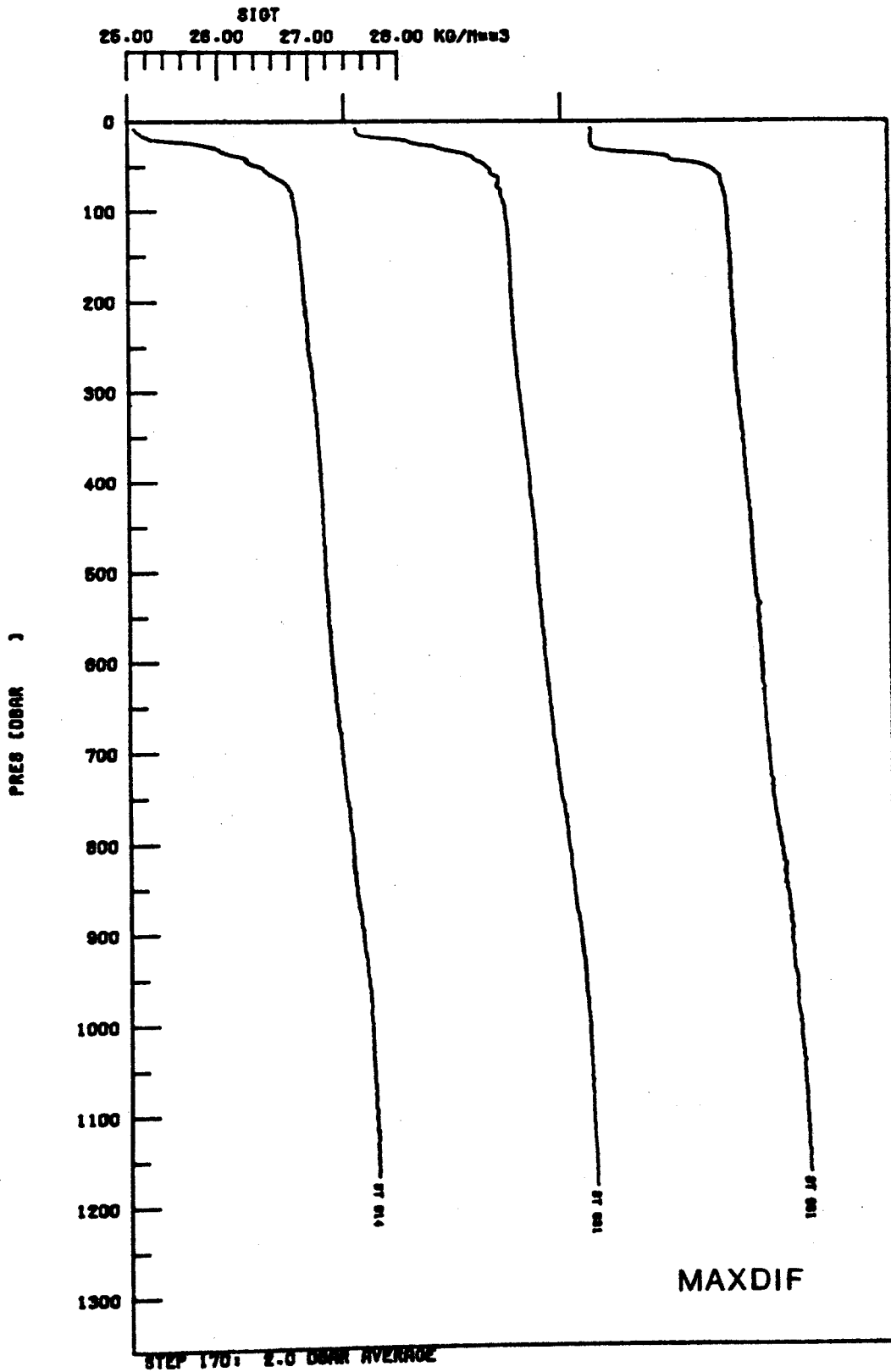
Fig. 20d
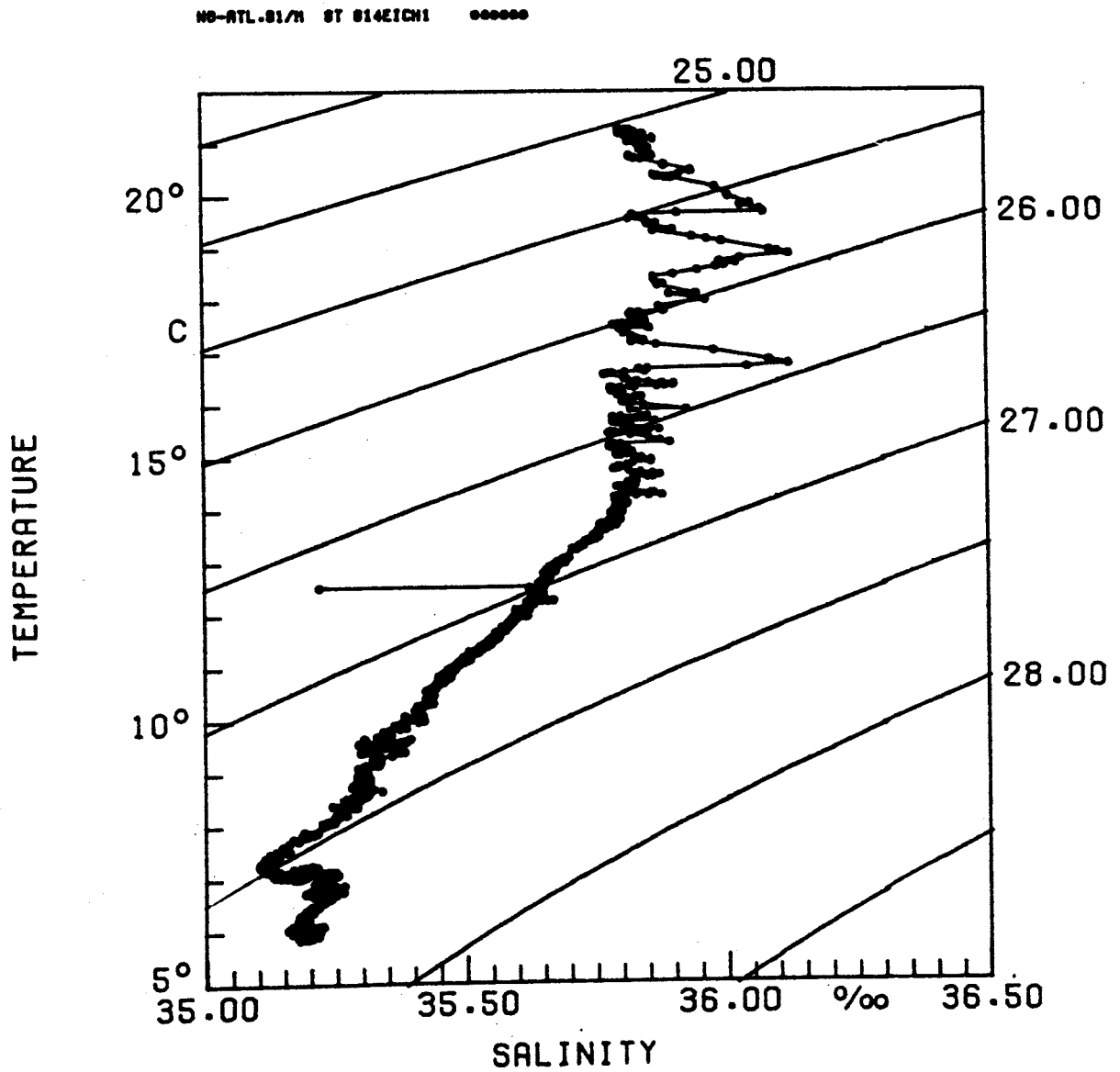
Fig. 21:   Station 614, raw data, TS-relation.
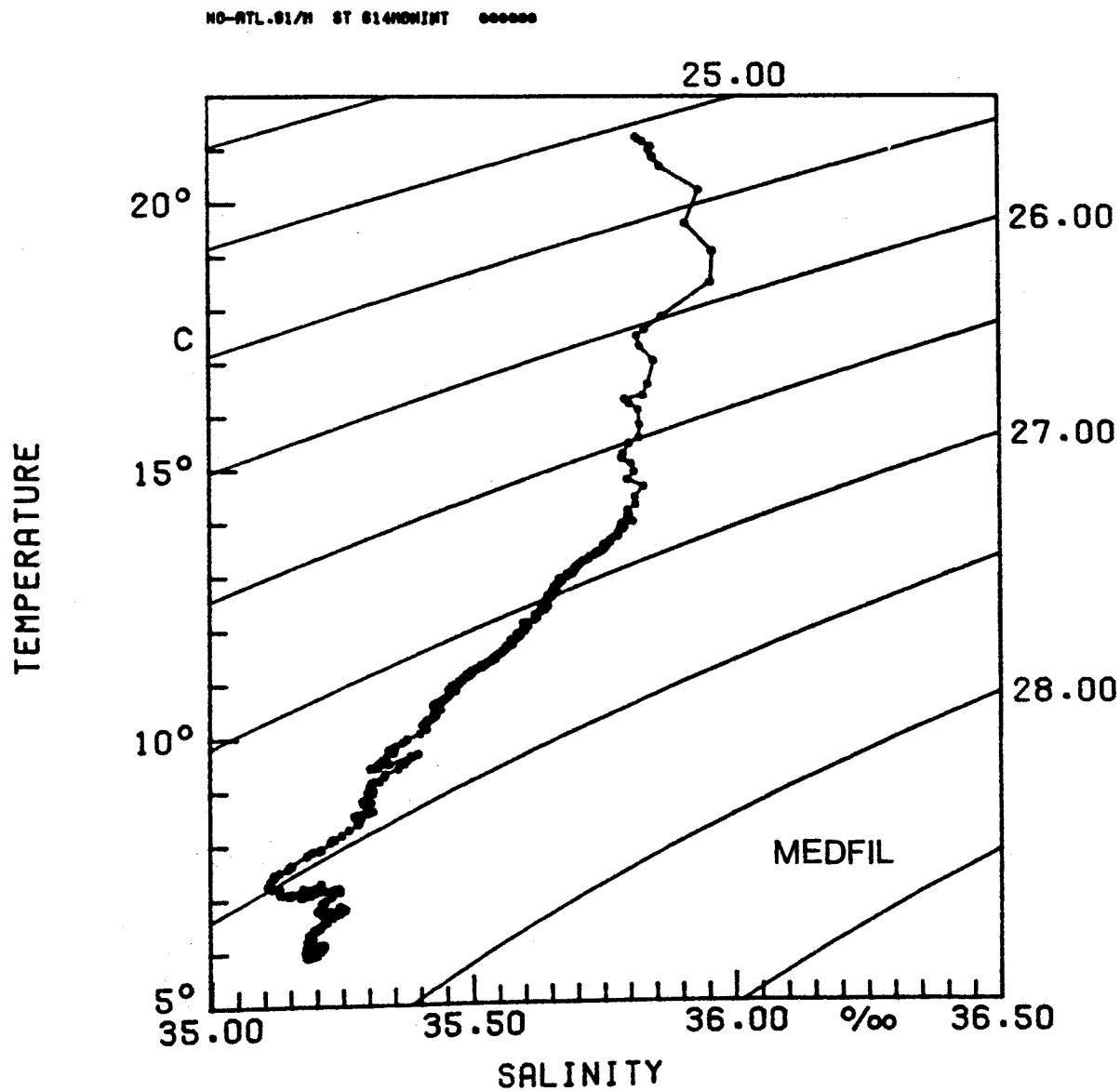
Fig. 22a-d: Station 614, final data, TS-relation for different editing techniques.
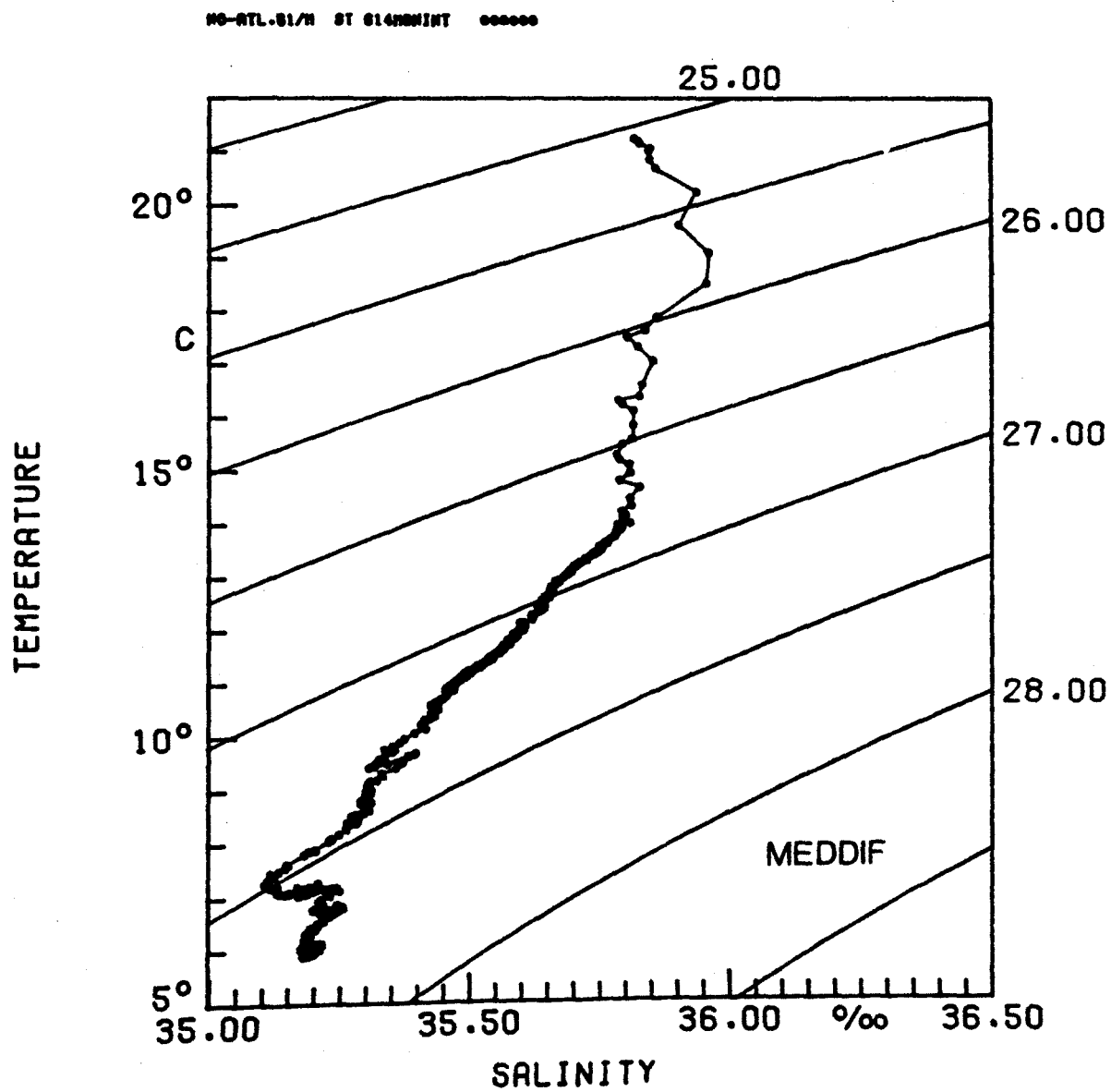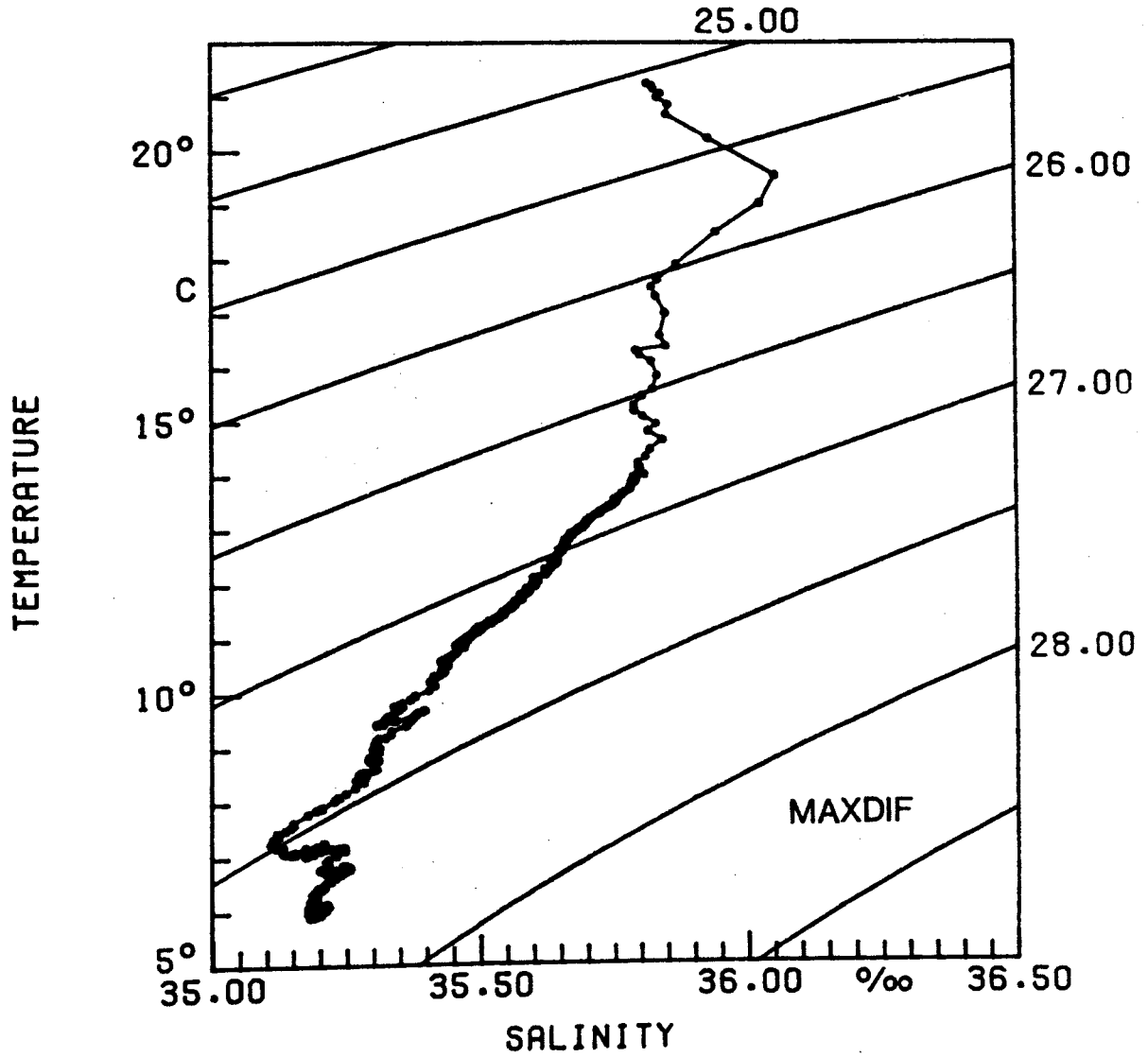
Fig. 22b
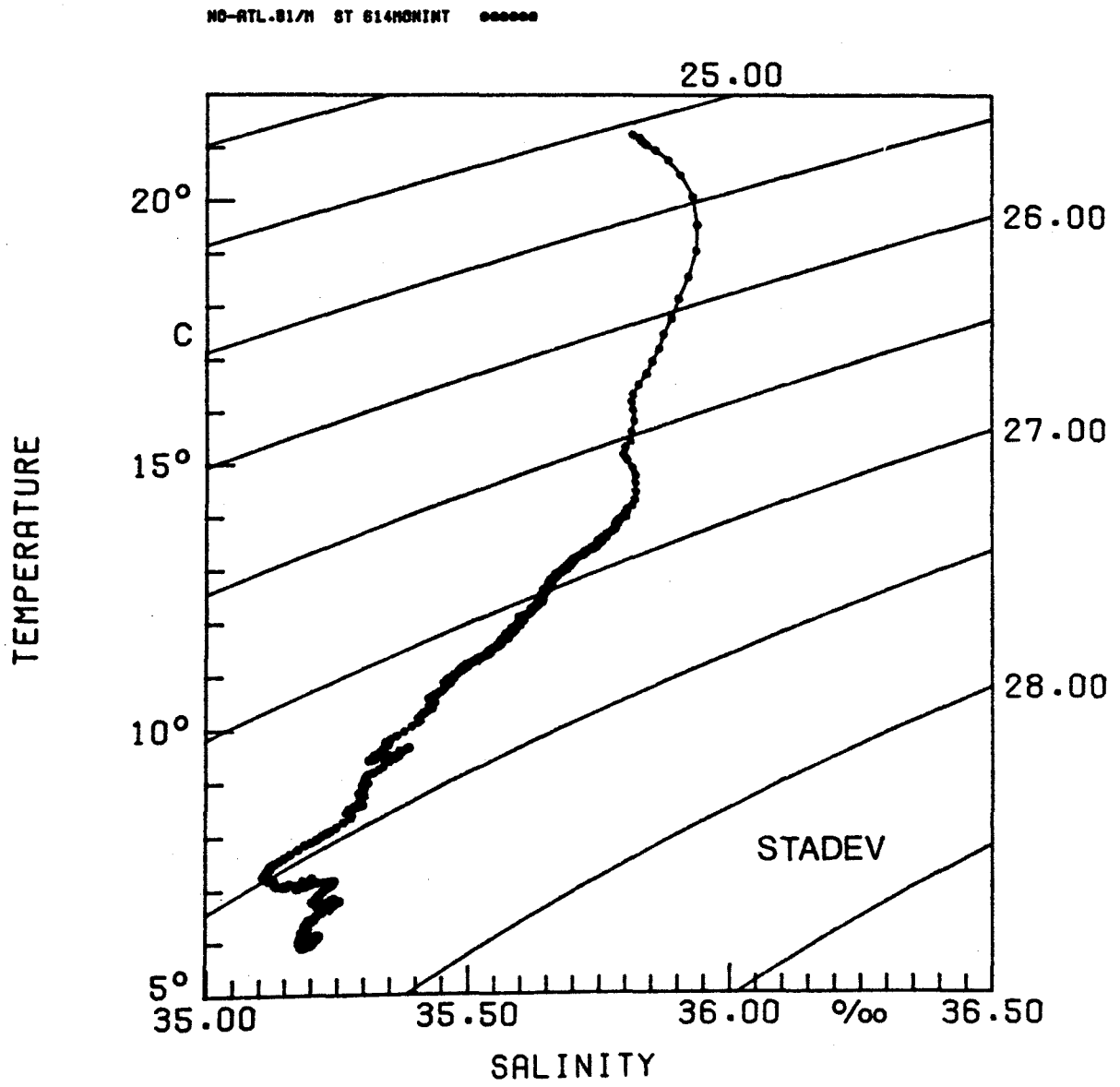
Fig. 22c

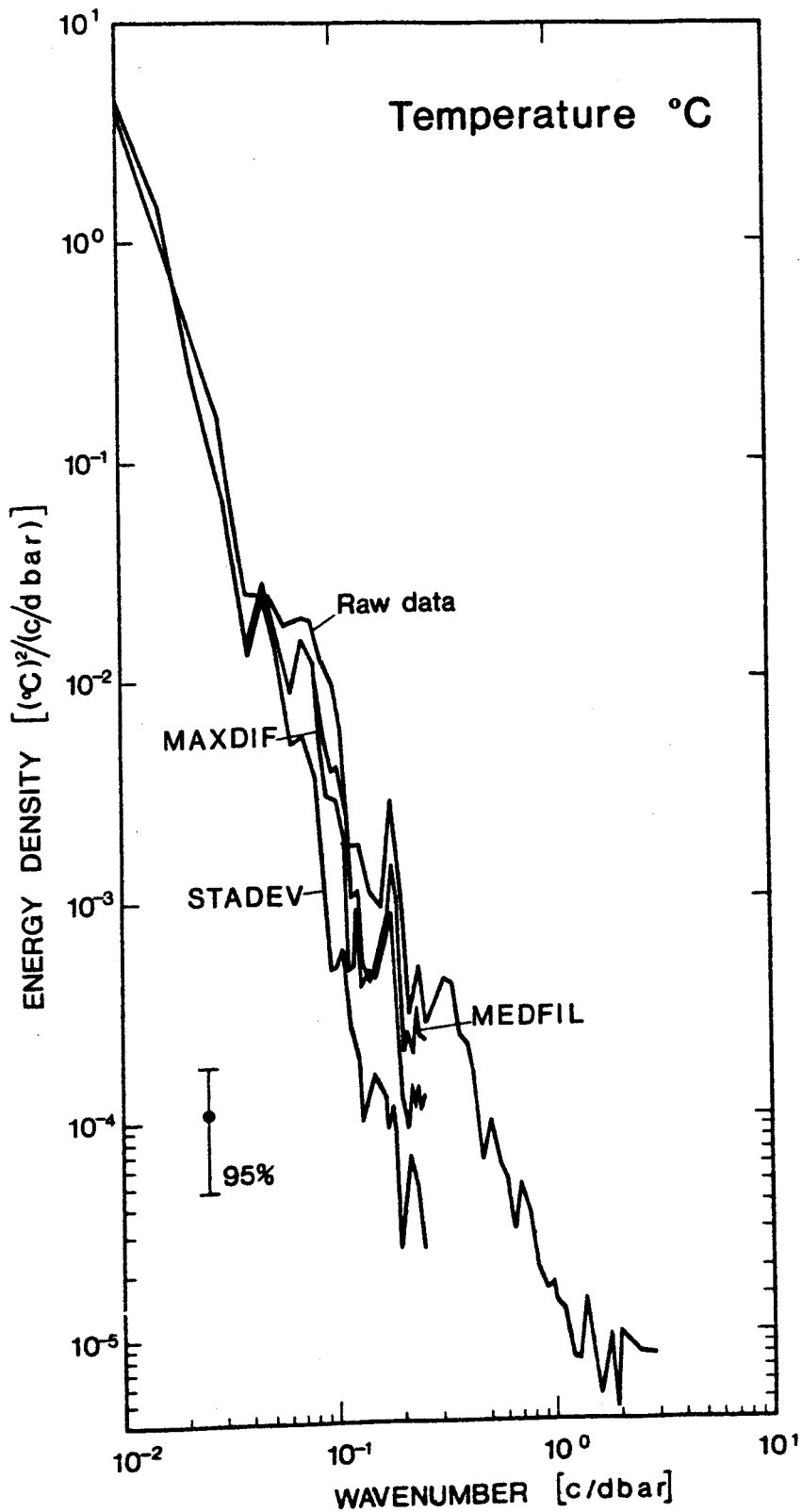NO-ATL.81/N ST 814MONINT 000000

Fig. 22d

Fig. 23a,b: Station 614: Vertical wavenumber spectra of (a) temperature and (b) salinity of raw data and final data for the editing techniques MEDFIL, MAXDIF and STADEV.
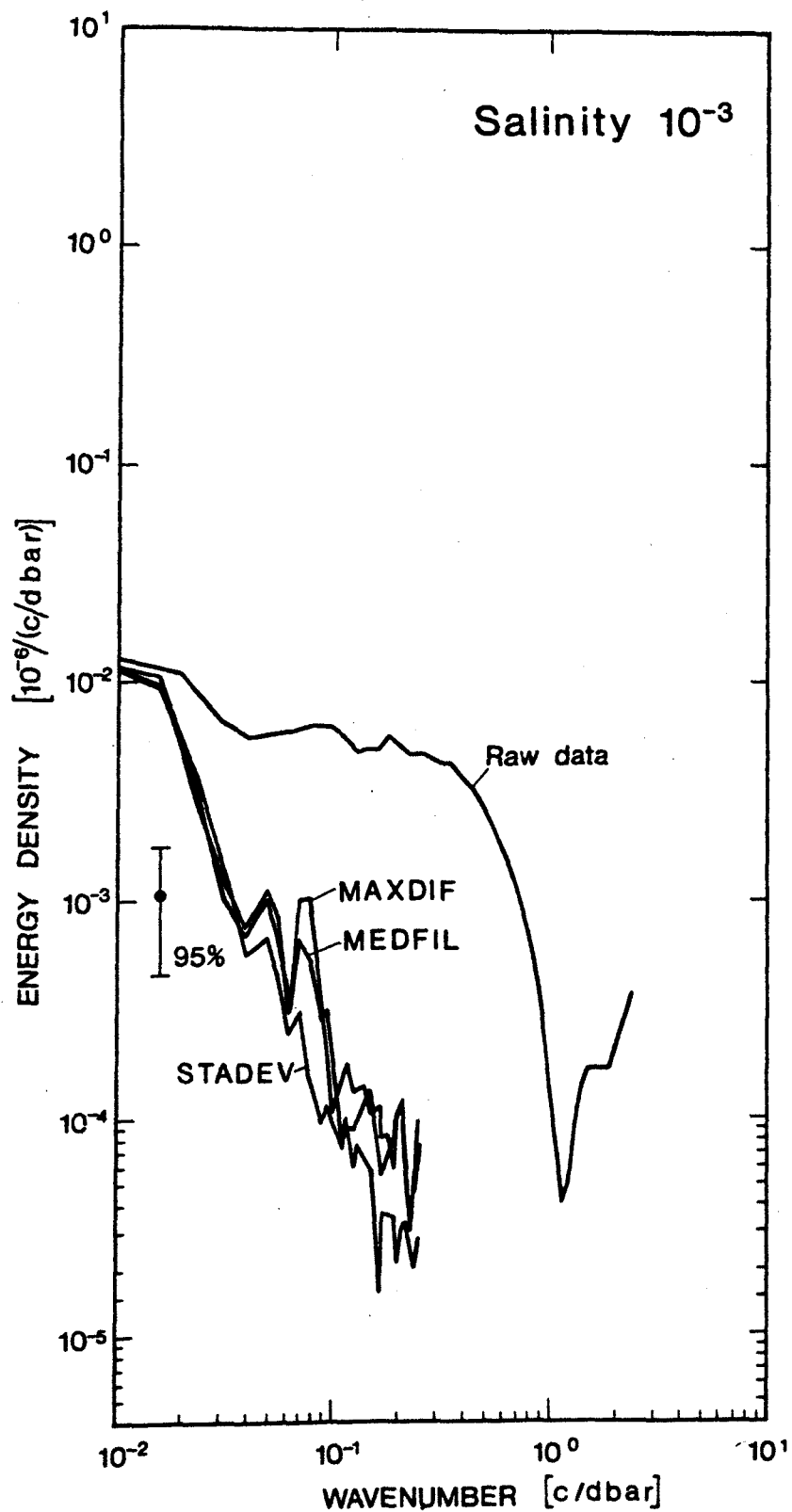
Fig. 23b