

## 大規模な海洋変動の研究に必要な水深 水温データの品質管理について

水野 恵 介\*

### Quality control of temperature-depth data for the studies on basin scale oceanographic variability

Keisuke MIZUNO

#### ABSTRACT

NRIFSF has deployed subsurface thermal field observing network in the tropical/subtropical area mainly in the Pacific Ocean by using 30-40 fisheries training vessels since mid 1960's. In order to manage these data, a semi-automated quality control (QC) system has been developed. In this paper, QC procedures relevant to our observation network which aims to study basin scale upper layer ocean variabilities are discussed. The spatial/temporal scale of interest is on the order of 100-1000km and a few months respectively. In this case, variance of signal is approximately the same as noise mainly due to smaller scale phenomenon (i. e. geophysical noise), and measurement error is smaller than mapping error. So, the main target of the QC is to detect time/location and temperature/depth errors caused by mis-reporting and malfunction of the equipment. For the temperature error detection, normalized temperature histogram for each 5°x 5° latitude/longitude grid by season and depth level is constructed. Most of histograms are found to be non-normal single mode distributions, indicating that ordinary 3-sigma data rejection method tends to reject correct data. Therefore, automatically searched doubtful data out of 3-sigma range are required to be re-checked by human inspection. The man-machine interactive QC system is described in the appendix. It is proposed that on-board automated QC system coupled with observing instruments should be developed.

#### 1. はじめに

近年、海洋が気候変動に重要な役割を果たすことが認識され、大洋規模の海洋変動に関する研究が注目されるようになった。さらに、気候変動と水産資源の変動との関係も盛んに論議されている (Glantz, 1992)。このため長期間にわたる大洋規模の観測データを収集し、データセットを構築して、海洋構造とその変動の特性を明らかにしようとする研究が盛んに行なわれている。このような研究では対象海域が

---

1995年1月9日受理 遠洋水産研究所業績 第316号

\* 遠洋水産研究所 (National Research Institute of Far Seas Fisheries; 7-1, Orido 5 chome, Shimizu-shi, SHIZUOKA, 424 JAPAN)

広いデータ収集には様々な観測実施機関の協力が不可欠であるとともに、個々の観測データの品質も保たねばならない。よって、観測やデータの収集・管理にあたる機関においては、データの品質管理が必要である。なぜなら、観測値は常にチェックして、その精度を把握していなければ使用に耐えないデータセットになるからである。

遠洋水産研究所では、太平洋およびインド洋で観測網を展開して、水深/水温データの収集・管理にあたっている。この観測資料は、主に表層（主温度躍層付近までの上層数100m）における大洋規模（空間スケール数100–1000Km）の季節・経年変動を抽出するために使用されており、この変動が、まぐろ類資源の分布・再生産に及ぼす影響評価や、変動自体の力学機構の解析に用いられている。実際、エル・ニーニョに代表される大洋規模の経年変動がその分布・再生産に影響することが見出されている（行縄ほか、1988）。さらに、このデータセットの統計的解析から、赤道からやや離れた緯度を西進するロスビー波とその西岸における反射が、エル・ニーニョの周期性に重要な役割を果たしていることが見出されている（Mizuno and White, 1990）。

当所におけるデータの品質管理は、人手によって行なわれてきたが、最近の国際共同研究などで観測が飛躍的に強化され、扱うデータが急増したため、旧来の手法で対応するのは困難となった。よって、著者らは品質管理の手法を探求し、収集したデータからできるだけ自動的に不良データを排除するシステムを作成した。本文は、大洋規模の海洋変動の研究を目的として観測網を展開する機関では、データの収集・管理のためどのようなデータの品質管理をするべきか論じたものである。また、実際に作成し、利用している品質管理システムを具体的に紹介するため、その概要を付録に掲げた。

## 2. 品質管理とは

本来、「品質管理」は大量生産される工業製品の製造工程で、これを目標の品質に維持するために実施される経営上の手法であって、広辞苑（1976）によれば、“製品の品質の安定化および向上を図ること。管理図法などの統計学的方法が用いられる。”とある。工業製品などの「品質管理」は不良品の排除のみでなく、より積極的な意味合いを持ち、個々の製造工程の検討・改良にフィードバックされる概念である。海洋観測データの場合、測定値の誤りや精度の悪いものを見出し、それらの原因を突き止めて、測定方法の検討や改良を行なってデータ精度の向上を目指すという意味では「品質管理」の概念と同じものであって、近年この語が使用されるようになってきている。

ただ、観測データは工業製品と違い、測定値の集まりであるとともに、その品質（精度）は研究目的に依存する。例えば、深層循環を研究するためには、深層での水温変動自体の幅（標準偏差）が $1/100^{\circ}\text{C}$ のオーダー（Roemmich and Wunsch, 1984）だから、これより悪い測定精度の資料は使えない。一方、表層の水温場の変動解析では、これほどの精度は必要としない。

データの品質管理の内容もデータの利用目的に依存する。Boyer and Levitus (1994) は、米国 NODC (National Oceanographic Data Center) に集められた水温、塩分および溶存酸素の歴史データ (Historical Data) を用いて、気候値（月別または季節別長期平均）データを作成する際に実施したデータの品質管理手法について論じている。この品質管理は、できるだけ正確な統計値を算出するための異常値の排除を目的としている。一方、できるだけ多くのデータを収集・整理して利用者の幅広いニーズに対応しなければならない“図書館”のような機能を果たすデータセンター（NODC や JODC など）では、収集されたデータは排除せずに保持し、以後の判断はユーザーに任せればよい。なぜなら、排除されたデータが、ある利用者には役立つような、“オーバーコントロール”の問題が生じるからである。実際、この

ようなデータセンターでは、疑わしいと判断した個々のデータにはフラグを付して、ユーザーに提供している。

データセンターが行うデータの集大成以前の段階でも、観測実施機関や、データ収集・管理機関など各段階においても何等かのデータの品質管理は行なわれているであろう。工業のような品質の維持・向上を目的とする品質管理を目標とするのであれば、データの品質管理は観測手法や測器の実状をできるだけ詳しく把握することが必要だから、調査活動の現場に近い所で実施することが望ましい。当研究所では様々な機関に依頼して大洋規模の表層（おおよそ主温度躍層まで）の水温分布を把握するための観測網を展開している。観測実施機関の間では、必ずしも測定手法や精度の統一が可能な訳ではないが、不良データの多い観測実施機関にフィードバックして手法や測器の改善を要求できる。よって、この場合の品質管理は、収集したデータから“悪いデータ”を排除するだけの受動的なものでなく、観測の目的に合致するようにデータの品質（精度）を維持・管理するためのものであり、工業の品質管理に近い。

### 3. 水深/水温データ収集体制

当所では1965年以来、太平洋および東部インド洋の熱帯・亜熱帯海域で海洋観測網を展開しており、このデータの収集・管理を業務として実施している。海洋観測は水産高校や水産試験場等の公庁船による“まぐろ資源調査”時に行なわれ、BT, DBT, CTD, XBT および各層観測による水深/水温観測が中心である。水深/水温データは所定深度(0, 10, 20, 30, 50, 75, 100, 125, 150, 175, 200, 250, 270, 300, 400, 500, 600, 700, 800, 1000m)の水温値として報告される(ただし、各層観測では観測深度の水温値)。図1に1965年から1986年までの太平洋熱帯海域における、水深/水温データの地理的分布を示した。これは、上記観測網の他、水産庁、米国スクリップス海洋研究所、JODC、その他の機関で収集されたデータを統合したもので、約200,000件の水深/水温データが含まれている。当所の観測網で得られたデータは、全体の40%近くを占めている (Mizuno and Yukinawa, 1991)。

この観測網に参加する公庁船の数は年により増減はあるが、30~40隻である。水温観測は以下の2つのモードを持つ。

①連続観測：日本から漁場に至る往路または復路で、航走6時間毎に実施する。

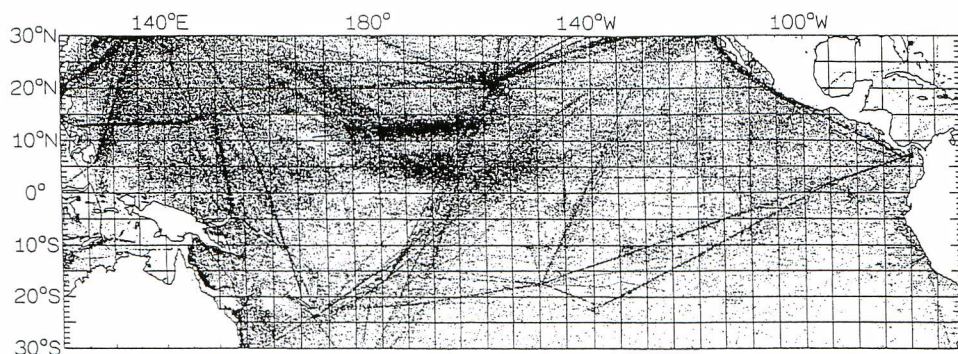


Fig. 1. Geophysical distribution of depth/temperature data in the tropical Pacific for the period of 1965-1986 (Mizuno and Yukinawa 1991).

②漁場観測：漁場に達してから、毎日1回実施する。

この観測網では、①では日本から赤道側へ放射状に延びる線上にデータが得られ、②では漁場に密集するデータが得られる。2つのモードの観測点群は図1に明瞭に現れている。データは航海終了後（原則として1ヶ月以内）、所定のデータシートまたはフロッピーディスクに入力されて当所に送付される。両者とも、個々の航海ごとに、計算機（ワークステーション）で処理可能な磁気媒体上の同一形式のファイル形式に統一され、品質管理を施した後、データベースに追加される。

#### 4. 必要とするデータの精度

第2節で、研究目的によりデータに必要な精度が異なることを述べたが、ここでは大規模な海洋変動を対象とする場合についてより具体的に論じてみよう。

この場合、対象とする変動の時空間スケールを明らかにしておく必要がある。いいかえれば、どの程度の“細かさ”で現象を見るかということである。例えば、時間スケール1ヶ月以上、空間スケール数100km以上の現象を対象とするならば、1ヶ月毎、数100km毎の平均像で表現され得る変動が求めたいシグナルである。一方、これより小さい時空間スケールの現象（例えば内部波や中規模渦など）はノイズとして処理すべきものである。これらは、実際の現象だから個々の観測値に反映されるが、大規模な海洋条件の解析には障害になるもので、“geophysical noise”（Meyers et al., 1991）と呼ばれている。一方、測定値は必ず測定誤差を伴う。これは、測器自身の誤差と、読み取り誤差や報告の際の誤りなどの人為誤差によるものである。よって、ノイズは“geophysical noise”と測定誤差の2つが原因である。図2は観測値とシグナルを模式的に示したものである。点線が求めたいシグナル、実線が観測値である。我々が知り得るのは実線で示された観測値のみであるが、これにはノイズが含まれている。よって、我々は何等かの方法により観測値からノイズの部分をできるだけ除去する操作を施して、シグナルを推定しているのである。

等値線解析を考えてみれば、一定の期間（月単位等）でデータをまとめ、さらに観測データは細かな構造を除去すべく空間的な平滑化処理（最適内挿法、デジタルフィルター、箱型平均など）を施して、大規模な現象を示す水温分布を描き出す。空間的に散らばった観測点のデータをもとに等値線を描き出す作業は、離散的なデータから、求めたいシグナルである大規模な水温分布を推定しているのである。等値線解析はこのシグナルをできるだけ正確に描き出そうとする作業に他ならない。

ここで注意しなければならないのは、ある観測点で、仮に測定誤差がゼロの測器で観測したとしても、この観測値には“geophysical noise”が含まれており、観測値をそのまま使用してもシグナルは得られないことである。すなわち、ノイズとみなすべき細かな構造を取り除き、目的とする大規模な変動をできるだけ正しく抽出する方法を用いなければならない。

最近、このような方法の1つとして最適内挿法（Gandin, 1963）が広く用いられるようになった。これは、ある点の推定値を周囲の観測値の重みつ

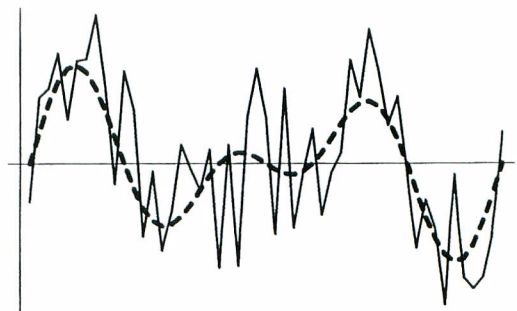


Fig. 2. Sketch of signal and noise. Broken line denotes a signal of interest, solid line shows observed data which include noise.

き平均によって推定する方法である。個々のデータの重みは過去のデータから得られた変動の性質を参照して、統計的に推定誤差を最も小さくするように決められる。実際、重みはシグナルとノイズの標準偏差の比（S/N比）と、水温場の相関関数（2点間の距離により相関がどう変化するかを表したもの）によって決まる。シグナルの推定精度は、同じ密度の観測であってもS/N比が小さい（ノイズが大きい）海域ほど悪くなる。太平洋の熱帯・亜熱帯海域の表層水温場のS/N比は0.5～1.0程度である（White and Bernstein, 1979; Meyers, et al., 1991）。この場合、シグナルの水温推定精度（内挿誤差）は、個々のデータの測定精度と観測密度に依存するが、現在達成されている測器の測定誤差からみて、観測密度に大きく依存する（White and Bernstein, 1979）。すなわち、離散的なデータから水温分布を推定するためには測定誤差の小さい高精度のデータがまばらにあるよりは、多少測定精度が劣っても観測密度が高い方がシグナルが精度良く求まることになる。

ちなみに、この方法による水温推定精度は熱帯・亜熱帯海域の表層水温の場合、場所や深度でやや異なるものの、表面水温でおおよそ0.25°C、それ以深では0.7°C程度と考えられている（Meyers et al., 1991）。これは、エル・ニーニョやラ・ニーニャのような大規模な変動は十分とらえられる精度である。個々の水温観測値に含まれる測定誤差はこの内挿誤差（0.7°C）と比べてできる限り小さくしなければならない。一方、深度に関しては、熱帯・亜熱帯海域で表層混合層の厚さの指標となる20°C等温面を最適内挿法でマッピングする場合の誤差は7 m程度（Meyers et al., 1991）と推定されているから、測定誤差はこれより小さくしなければならない。

## 5. 水深／水温データの測定誤差

ある観測点における観測データは多くのデータ項目から成るが、水温分布を描き出すために必要なデータ項目は基本的には、測定時刻、測定場所（緯経度）および各深度における水温の観測値である。すなわち、水深／水温データは、ある場所、ある深さ、ある時刻の水温のデータの集りとしてとらえることができる。すなわち、

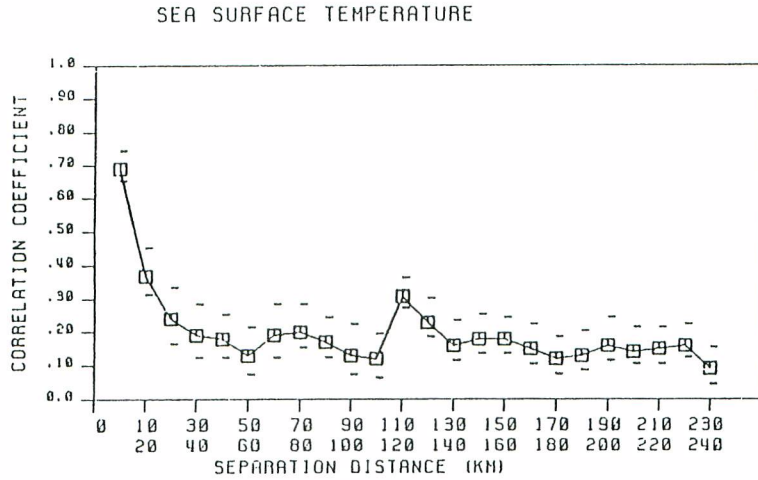
$$T = F(x, y, p, t) \quad (1)$$

ここに、 $T$ は水温、 $x$ は緯度、 $y$ は経度、 $p$ は圧力、 $t$ は時刻で、いずれも測定値である。ここでは、前節で説明した“geophysical noise”は考慮しないものとする。いずれの測定値にも測器自体の誤差と人為誤差が含まれる。以下に各々の観測項目について論じる。

### 1) 緯経度（ $x, y$ ）および時刻（ $t$ ）の誤差

緯経度（ $x, y$ ）は航海計器（この観測網ではロランC、GPSが主体）で測定され、その位置測定精度は、ロランCで100～500m、GPSで10～30m程度のオーダー（大日本水産会、1990）である。空間スケール数100kmの構造をシグナルとしてマッピングする立場からは、これより2桁以上小さい誤差は問題にならない。また時間的には月毎の平均像をもとに月毎の変化を対象としているから、時刻（ $t$ ）に関しても通常の時計で全く問題はない。

問題なのは人為誤差で、人が航海計器や時計を読んでデータを作成するから、緯経度の数値の読み誤りや、東・西経または南・北緯の取り違いがしばしば起こり得る。図3は 既往の海洋気象データが集められたCOADS データセット（Release 1）を用いて、熱帯域における表面水温の任意の2地点間の相関を距離の関数として計算したものである（Lander and Morrissey, 1987）。この図によれば、相



**Fig. 3.** Autocorrelation function in terms of space lag (in km) of sea surface temperature along 150E and 10S-10N range. Data are archived in the COADS data set (Lander M. A. and M.L. Morrissey 1987).

**Table 1.** Types and models of equipments with their accuracy used for the depth/temperature observation network of Far Seas Lab.

種 類	型 式	メーカー	水温精度	深度精度
機械式 BT	TS-US Type 3	鶴見精機	0.2	3% FS *
デジタル式 BT	DBT MOX-BT 2F	環境計測システム	0.05	0.3% FS
	DBT-V	環境計測システム	0.05	0.25% FS
	マイコン BT	鶴見精機	0.05	0.3% FS
	ATD1000	アレック電子	0.02	0.11% FS
	MBT-600	村山電機	0.1	< 1% FS
XBT	MK-2S	鶴見精機	本文参照	本文参照
	MK-30	鶴見精機	本文参照	本文参照
CTD	Mk3b	Neil Brown	0.005	0.1% FS
	AST1000	アレック電子	0.02	0.11% FS
	マイコン-STD	鶴見精機	0.05	0.3% FS
	SBE19	Sea Bird	0.01	0.5% FS
転倒温度計			0.01	

\* FS ; Full Scale

種類, 型式, メーカーについては, 1993年に遠水研が観測実施機関に行なったアンケート結果(水野 1994)を参照。精度については各社のカタログを参照した。

関は全般的には距離が離れるほど低くなる傾向にあるが、110~120kmの範囲で有意に高くなっている。これは実際には考えにくい現象であり、緯度1°の間隔で似た温度が報告されているというのは位置が緯度で1°誤った報告の数がかなりの数にのぼっていることを示している。年月日、時刻に関しても報告の誤りがしばしば見られる。

## 2) 水温 ( $T$ ) 及び圧力 ( $p$ ) の誤差

水温及び圧力の測定にはCTD, XBT, 転倒温度計などが用いられる。当所の水深/水温観測網で使用されている測器を表1に示した。ただし、機械式BTでは、ほとんど使用されなくなっている。これらの測器の水温測定誤差は、XBTを除けば0.1°C以下であり、内挿誤差よりはほぼ1桁小さくほとんど問題にならない。XBTの水温測定精度は、様々な報告があるが (Flierl and Robinson, 1977; Hanawa and Yoritaka, 1987; Heinmiller et al., 1983 など), いずれも0.2°C以下であり、少なくとも内挿誤差よりは小さい。

深度に関しては、デジタル式BTおよびCTDの圧力センサーは0.1~0.5%程度の誤差がある。ここで対象とする主温度躍層までの上層数100mであれば、この程度の誤差は内挿誤差よりかなり小さく、測定誤差はあまり問題にならない。

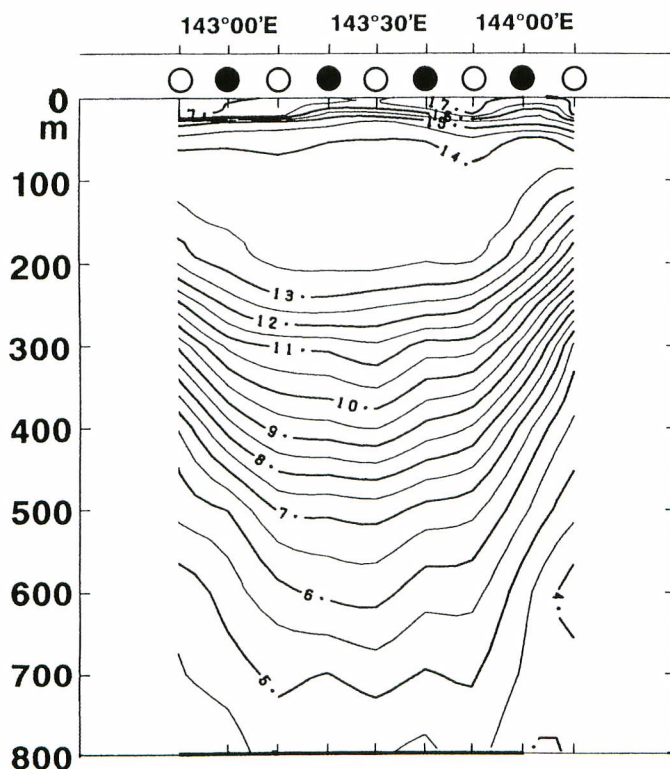


Fig. 4. An example of "XBT wave". XBT (solid circle) and CTD (open circle) observation were conducted alternately along an observation line. This section transects a warm core ring along 38°15'N off the east coast of Japan in May 1995 by Wakataka Maru.

XBT の深度（この場合は圧力でなく距離で定義される）はプローブ着水後の経過時間から換算される。この時間－深度換算式は実験的に求められた 2 次関数である。しかし、CTD と併用して同じ観測線上で交互に観測して、鉛直断面図を作成すると、等温線が 1 測点おきに波打つ“XBT ウェーブ”を生じる（図 4）。T-4, T-6 (460m 用) および T-7 (760m 用) プローブについてメーカーから供給された換算式に問題が指摘され、より良い精度の式が求められている (Hallock and Teague, 1992; Hanawa and Yoritaka, 1987)。1996 年 1 月より国際的に統一して新しい補正式を用いて既往の XBT データを再計算することになっている (WMO, 1994)。既存の式と新しい補正式について経過時間と到達深度の関係を図 5 に示した。これによれば、既存の式では 200m では 5 m を越える誤差を生じる。これは最適内挿法で 20°C 等温面深度 (熱帯ではおおよそ 100–200m) を推定する場合の内挿誤差 (7 m 程度) と同等な大きさになってしまうため、新しい式を使って補正することが望ましい。ただし、データセンターに報告するデータは、新旧の式が混在しては混乱を生ずるので、1996 年 1 月以前は全て旧の式で報告し、以後は一斉に新しい式を用いることになっている。

各層観測では、深度方向に離散的にデータが採取される。転倒温度計の温度精度は十分 ( $1/100^{\circ}\text{C}$ ) であるが、深度測定については通常 200m までは傾角で決定するため、これが大きくなるとワイヤーの形状が直線とみなせなくなるので精度に問題が生じる。それ以深では被圧・防圧温度計の温度差から深度を測定するが、この方法では読み取り誤差が各々  $1/100^{\circ}\text{C}$  程度で、これは約 1 m の圧力差に相当するから、深度誤差は 1–2 m の誤差に収まるはずである。ただし、ワイヤーが流れなどのために水中で安定せずに測定した場合には水温計が安定せずこれより大きな誤差を生じる。また、途中転倒などで精度の悪いデータがしばしば見られる。

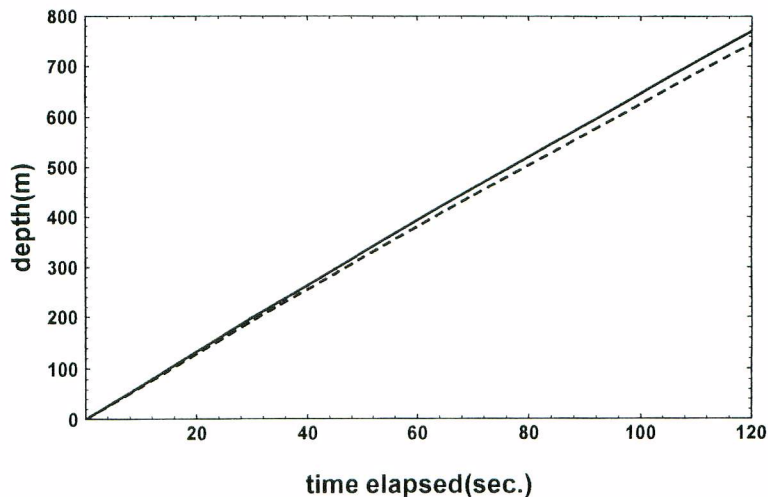


Fig. 5. XBT probe depth (m) in terms of elapsed time since drop into sea. Broken line; manufacturer's fall rate equation ( $d=6.472t-0.00216 t^2$ ). Solid line; new fall rate equation ( $d=6.691t-0.00225 t^2$ ).



## 6. 品質管理の種類

前節の論議から、大規模な表層の水温分布の把握のためには水深／水温観測データは位置、時刻、水深、水温のいずれも測器が正常ならば測定精度自体にはほぼ問題がない。ただし、XBTの水深については新たな換算式に変換する補正が必要である。人為誤差については、いずれの項目にも含まれる可能性があるとともに、測器自体が不調で誤ったデータを得る場合もあり、全ての項目に検査が必要である。よって、品質管理には以下の検査が必要となろう。

- 1) データ形式の整合性検査
- 2) 同一データの多重登録の検査
- 3) 時刻／位置の検査
- 4) 深度／水温の検査

上記の1)はデータが規格の形式に合致しているかどうかの検査である。レコード長や入力項目の位置などの形式上の問題点と、各項目の値も検査する。観測項目が、緯経度、日付、水温などのような数値であればそれらのあり得るべき値の範囲を定めてこれと比較する(例えば水温ならば $-2^{\circ}\text{C}$ ~ $32^{\circ}\text{C}$ )。測器名などの文字型のデータは、あり得るべき全ての場合を網羅したリストを作成して、データはその中に含まれているかどうかを検査する。

最近の傾向として測定データは現場で自動的に磁気媒体に記録されるタイプの測器が普及しつつある。ただし、その記録形式は測器のメーカーや型式により異なるため、現場で取得されたデータを報告する際の所定の形式に変換することが必要である。この変換が正確に行なわれていれば形式上の問題は少なくなっていくであろう。

上記の2)の多重登録は、1隻の船の観測結果を複数の機関から報告を受けたり、同一機関から報告が繰り返される等の理由で、しばしば起こる問題である。ある海域の物理量の分布を表示する場合、通常は空間的に離散的に分布するデータの平均化操作を行なうので、多重登録されたデータはそのデータの重みが不当に大きくなり、偏りのない推定値が得られないので解析の障害となる。多重登録されたデータが全く同一であれば機械的に検出・削除することは容易である。しかし、実際には同一のデータを取得直後に報告し、その後補正を加えて報告する機関も多く、船名、時刻、位置を検索のキーとして、同一なものを検出し、観測機関に最終のデータを問い合せて、それ以外は削除する。不明の場合は、最も新しく報告されたものを採用する。

上記の1)及び2)は定型的な作業であり、計算機を用いれば、ほぼ誤りなく、短時間で検査できる性質のものである。一方、3)および4)の問題は完全な解決方法はないが、ある程度は対処できる。

上記の3)に関しては前節で述べたように、位置の誤りは、しばしば起こるものである。この検査方法は、まず同じ船の報告を時間順に並べ、隣接する2点間の距離を時間差で割って船速を求める。このため、データには船名コードが不可欠である。この船速が通常の船の航行速度よりも非常に大きければ、時刻、位置、または船名に誤りがあると考えられる。この速度を“基準速度”と呼ぶことにする。調査船の場合、基準速度は15ノットを上回るのは稀であろう。この検査にかかったデータは観測機関に問い合せて修正するが、実際には正確な記録にさかのぼれず修正が困難な場合が多い。しかし、明らかに推定できる場合もあり、この場合は東経・西経の誤り、度数の誤り、日付の誤りが原因であることが多く、稀には、船名コードの誤りもある。将来的には、船の航法装置が観測機器とつながれて位置情報が自動的にデータに取り込まれるようになると考えられ、次第にこの問題は解決されるであろう。

上記の4)の深度/水温の検査は、1)～3)のような明快な判断が難しい作業である。この検査方法は、既往データを標本として、その統計的性質を把握し、新たに得られた個々のデータと比較して、あまり標本とかけ離れていれば異常データと判断するものである。ただし、個々のデータは上記の1)～3)の検査を経たものを用いる。

標本データセットの作成方法や、異常値の判断方法などは、決定的なものは無く、データ量やその統計的性質から判断されるものである。それらの検討は次節以下で行なうことにする。

## 7. 標本データセットの作成

標本データセットを作成するために用いる既往のデータには精度の悪いものや、誤りのあるものが含まれている可能性があり、まずこれを除去しなければならない。これを除去した既往データの集りを標本データセットとする。さらに統計的性質は、海域や季節によって異なるから、時間・空間的に一定の範囲で1つの標本としなければ実用的ではない。品質管理を行うデータは大洋規模の海洋変動(時空間スケール、数100km以上、1ヶ月以上)の解析に使用されるから、1つの標本の時空間範囲はこれと同程度にすることが望ましい。ただし、標本から正確な統計値を推定するためにはデータ数が多くなければならないから、実際にはデータ数との妥協とならざるを得ない。ここでは、図1に示されたデータについて、空間的には緯経度 $5^{\circ} \times 5^{\circ}$ の区画毎にまとめ、時間的には各区画につき季節毎に(1-3月, 4-6月, 7-9月, 10-12月)分けて、標準層毎の標本を作成した。ただし、対象海域は $30^{\circ} \text{S} \sim 30^{\circ} \text{N}$ ;  $120^{\circ} \text{E} \sim 70^{\circ} \text{W}$ の太平洋熱帯域である。用いた既往のデータは1965年から1986年までの20年間のものであるから、熱帯域の主要な変動であるエル・ニーニョは複数含まれており、この中には今世紀最大と言われた1982-1983年のものもある。この標本を参照すれば、将来これを凌ぐエル・ニーニョが出現しない限り正しいデータを異常値として排除することはない。

標本データセットの作成には3段階の異常データの除去処理を行なった。すなわち、第1に各測点の鉛直プロファイルを手作業で目視で検査した。すなわち、各区画の季節毎のデータの鉛直プロファイルを一度に表示し、異常と思われるプロファイルを取り除く。除去されたもののほとんどは、報告の誤りと思われる大きな水温の逆転、XBTプローブの着底やワイヤーの途中切断・漏電による異常なプロファイルである。

第2に各区画で季節毎に所定層毎の平均値と標準偏差を計算する。個々の水温プロファイルにおいて平均値からのずれが標準偏差の3倍を越える水温が所定深度の1ヶ所にでもあれば、この水温プロファイルはデータセットから除去する。この操作を除去されるデータが無くなるまで繰り返した。

最後に、こうして得られたデータセットを更に第1段階で行なったのと同様に個々のプロファイルを目視で検査して、異常なデータを判断して除き最終的なデータセットとした。

Mizuno and Yukinawa (1991)は太平洋熱帯域の水深/水温データについて以上の方法で異常値の除去を行なったところ、最も多かった区画で22回の繰り返し計算を要し、全体で最初のデータセットの9.5%の水温プロファイルが除去された。これは最大/最小値付近のデータを各々5%程度除去したことになる。この方法では、中規模渦などの実際の海洋現象も除去した可能性は否定できない。しかし、我々の見たいのは渦などの細かい構造は平滑化された大規模な水温分布であり、仮に実際のデータが除去されたとしても標本データの統計的な性質に大きな影響を与えるとは思われない。

以後これを参照すべき標本データセットとして品質管理に用いる。Mizuno and Yukinawa (1991)はこれをもとに各区画・各季節毎に、所定深度の平均値により水温鉛直プロファイルを作成した(図6)。標準偏差は水温鉛直プロファイルの所定深度毎に横棒で示してある。

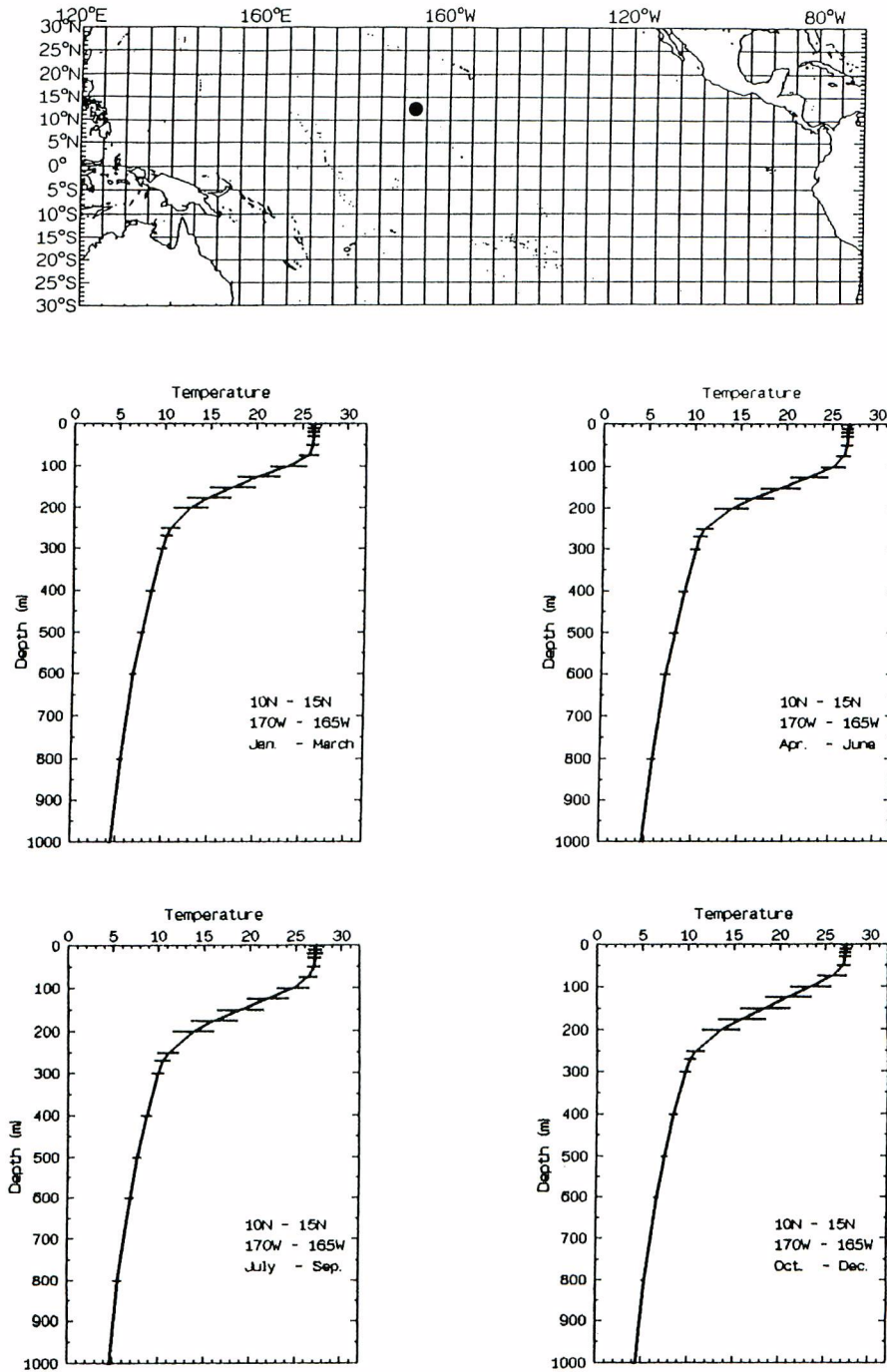


Fig. 6. An example of seasonal vertical temperature profile with standard deviation range for 5x5 latitude/longitude square. The upper panel shows the location of the square.

## 8. 水温ヒストグラムの検討

標本データの水温ヒストグラムは、ある海域のある時期にどのような水温が現われやすい傾向にあるかを示すものだから、品質管理において正常・異常値の判断にとって、ヒストグラムの形状は重要である。すなわち、分布型が正規分布でなく、歪みが大きい場合は、仮に $3\sigma$ 等を取捨選択の基準とすれば、正しいデータを除去する可能性が強い。また、双峰型でこの中間の値を取る頻度が非常に小さいような場合には、実際にはあまり起こり得ない中間的なデータを正常と判断してしまう。よって、各区画で頻度分布の検討が必要となる。また、分布型には該当の海域の海洋変動の特性を示唆することもある。

### (1) ヒストグラムの作成

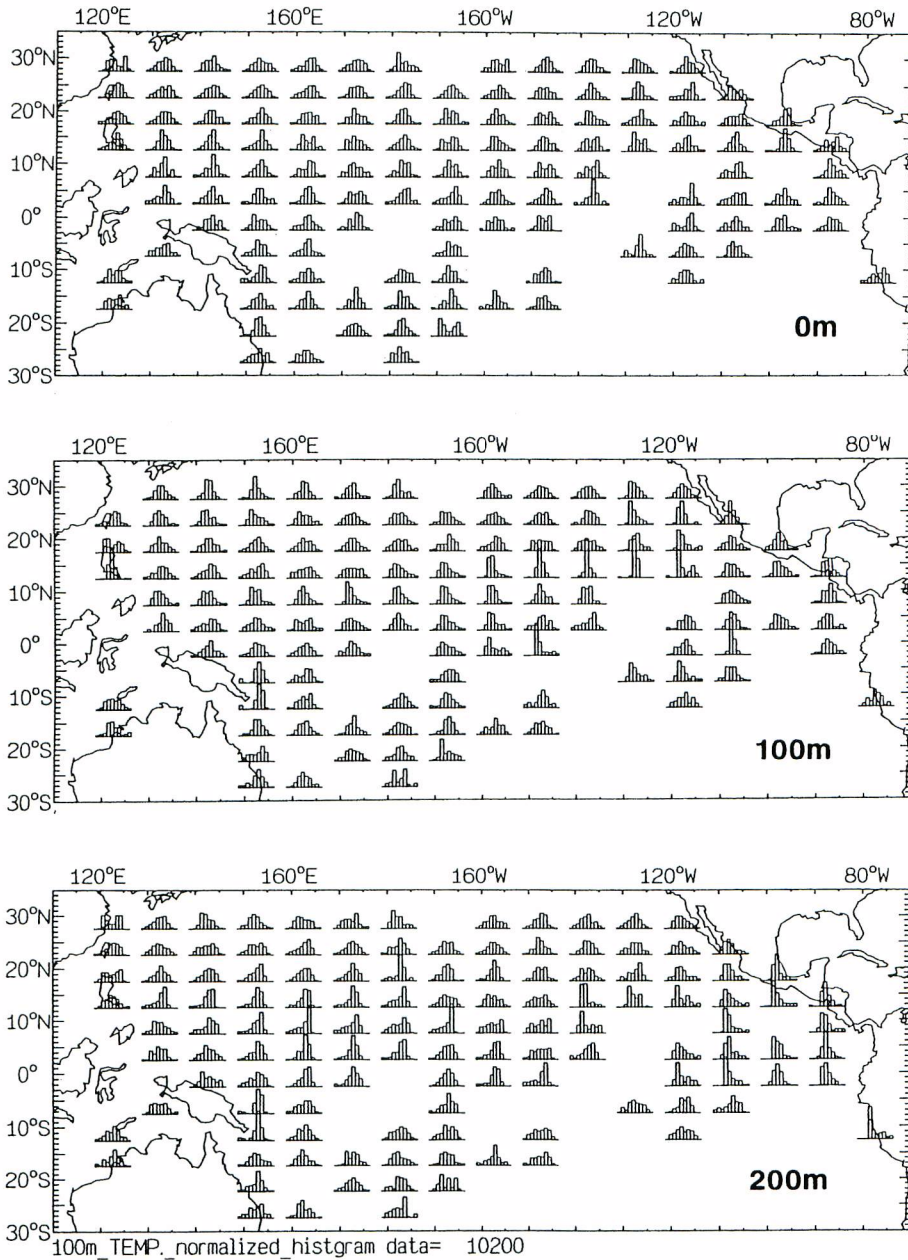
各々の標本のヒストグラムを作成して、その分布型を統計的に検討するには十分なデータ数が必要である。ここでは、鈴木(1968)に従って、データ数が50以下のものは対象としなかった。各標本は地理的位置、季節、深度が異なるため、取り得る水温値の範囲は標本毎に大きく異なる。よって、水温値は正規化したものを使用する。すなわち、標本毎に平均および標準偏差を計算し、1つのサンプルの値 $X$ は平均からの偏差を標準偏差で除した値 $x$ に変換する。この正規化した値でヒストグラムを作成した。

ヒストグラムの作成方法は古くから論議されてきたもので、階級の取り方については、各標本の標本数が50~1000程度であるので、Sturges(1926)の等間隔階級数の基準に従って、階級数を

$$K = 1 + \log N / \log 2 \quad (2)$$

とすれば、6~11階級必要である。正規分布のパーセント点を目安として5%点(両側)は $1.96 \div 2$ で、多くのデータは $|x| < 2$ の範囲に入る。また、分布形の適合度を検討する $\chi^2$ 検定では1つの階級の最小の度数は少なくとも5が必要となる。 $|x| < 2$ の範囲で6~11階級となればよいので、区切り良く階級間隔を0.5とし、上記範囲では8階級とした。

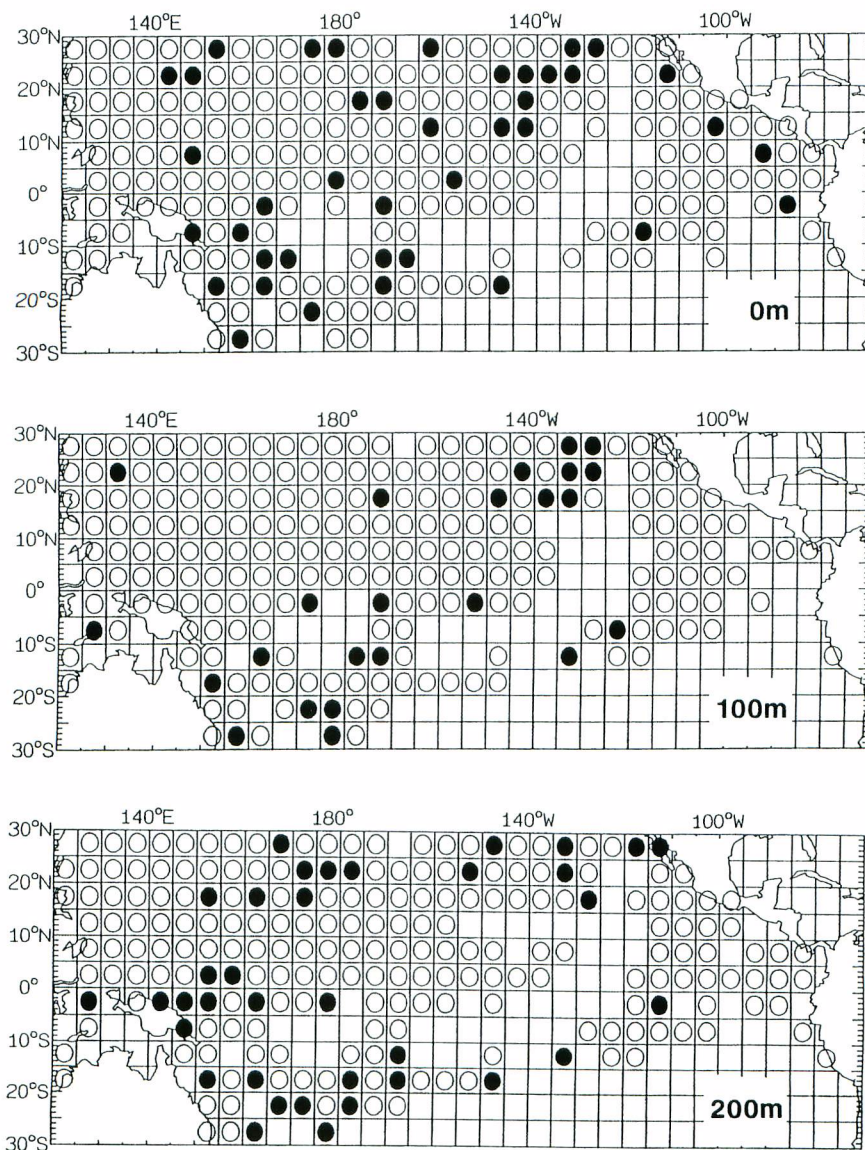
表面、100m、200m深における1~3月の度数分布を図7に示す。表面水温(同図上)では $15^\circ\text{N}$ 以北では沿岸部をのぞき、単峰型がほとんどで、平均値の左右で比較対称な分布を示す。しかし、赤道付近では分布型に対称性を欠くものが多い。100m深でも $15^\circ\text{N}$ 以北は単峰型で、対称な分布を示す傾向が見られ、赤道付近から $15^\circ\text{N}$ までは分布に歪みのあるものが多い。赤道付近の緯度帯( $5^\circ\text{S} \sim 15^\circ\text{N}$ )では、東西でヒストグラムの分布型に違いがあり、西側では平均値より高温側にモードが来ることが多く、東側では低温側にモードがある。この中間の $140 \sim 150^\circ\text{W}$ 付近では分布が一様でモードが明瞭でない。西(東)側では高(低)温を基調とし、でしばしば低(高)温になることを表わしている。赤道域の東側の表層には通常100mを超す厚い混合層があり、西側で薄くなる。エル・ニーニョ時にはこの厚い表層の暖水(混合層)が東に移動するため上記のヒストグラムが現れると考えられる。ただし、表面のヒストグラムでは、同様な傾向はあまりはつきりせず、表面水温は暖水移動のみでなく、大気との熱交換の諸過程も含まれるより複雑な過程で決まることを示していると思われる。200m深では、表層の混合層下面よりも一般に深く、必ずしもエル・ニーニョが反映されるものではない。ヒストグラムも100mのそれとかなり様相が異なっている。赤道域では、単峰型で対称なものが比較的多くなっている。しかし、北赤道反流域にあたる $5 \sim 10^\circ\text{N}$ 帯はヒストグラムには歪みの大きいものが目立つ。この海流の位置や流量の変化は、エル・ニーニョと密接に関連しており(Wyrtki, 1975)、この変化は躍層深度に影響を及ぼすため、このようなヒストグラムが現れると思われる。



**Fig. 7.** Normalized histograms of temperature observed in the 5°x5° latitude/longitude squares during January - March, but are shown each 10° zonally. Upper, middle and lower panel shows the histogram for 0m, 100m, 200m respectively. In each histogram, eight classes centered at the mean are shown, and the size of the bin is half of the standard deviation. Scale of the frequency is that the length of 5° in latitude (to the bottom of the upper histogram) coincides with 42%.

(2) 分布形の特徴 ( $\chi^2$ 検定)

上記の様に、水温のヒストグラムは海域・深度により形状は異なるが、全体としては各深度で単峰型がほとんどであった。しかし、多くの場合、正規分布としてはかなり歪んでいるように見える。そこで各区画のヒストグラムが正規分布とみなせるか否かを  $\chi^2$ 検定で調べてみた。ただし、奥野(1986)に従っ



**Fig. 8.** Results of  $\chi^2$  test (95% confidence level) for the shape of histogram distribution under the normal distribution hypothesis. Open (solid) circle denotes the histogram which discards (holds) normal distribution hypothesis. The histograms are the same in the previous figure.

て、度数の記録が小さいものは近隣の階級にまとめ、得られた階級の数 $5$ 未満の場合は、対象としなかった。結果を、 $1-3$ 月の表面、 $100\text{m}$ 、 $200\text{m}$ 深について図 $8$ に示す。何れも、黒丸で示した正規分布とみなせるもの(有意水準 $95\%$ )は少なく、 $10-15\%$ に過ぎない。この結果は他の季節や $1000\text{m}$ 深までの他の所定深度でもほぼ同じ傾向を示し(図は略)、ヒストグラムは正規分布の仮説を棄却するものが多数であった。よって、全体的には水温のヒストグラムは正規分布とはみなせず、単峰型ではあるが左または右に歪んだ分布が一般的である。

## 9. 水深／水温データの検査方法

報告された水深／水温観測データを検査する際には、深度は正しいものとして、その深度における水温が妥当か否かを判定する方法をとる。これは、深度に誤りがあれば水温値が異常値として検出されることを期待するものである。

水温の検査では、報告された深度における平均値と標準偏差を求め、観測データと平均値との距離を標準偏差を尺度にして計り、これが余り大きければ(標準偏差の $3$ 倍以上など)、問題ありとみなす方法が一般的である。ただし、前節で検討した標本データセットでは海域・季節に分けた大部分の標本は単峰型ではあるが正規分布としては歪みが大きい分布型なので、 $3\sigma$ 以上のものは正規分布の場合よりも多くなるから、正常なデータも排除する傾向が大きくなる。

一方、実際の標本のヒストグラムから累積度数分布を作り、その両端にあたる稀にしか起こり得ない事象は異常とみなす方法もある。これは、例えば累積確率の下(上)側で $0.5\%$ ( $99.5\%$ )などと定め、対応するパーセント点の範囲の外側の観測値は異常値としてみなす方法で、分布型によらず利用できる。この他、標本の最大／最小値を判断基準とし、この範囲を越えたデータは異常とみなす方法もある。これだと、過去最高または最低の値は全て問題ありと判断される。

何れの方法も、統計的な比較に過ぎないので、正確な判定を期待できるものではない。よって、“疑わしいもの”は異常値として指摘しておいて、最終的には人間が判断すべきである。このような方針をとれば、その基準はやや厳しいものにしておく方が良い。よって、上記の平均から $3\sigma$ の範囲か、累積度数分布における一定のパーセント点の範囲( $0.5\%$ と $99.5\%$ など)の何れかを採用するのが実用であろう。我々の場合、計算が簡単な前者の方法を使用している。また判断にできるだけ多くの情報を加味するため、隣接する観測点の水温プロファイルを並べて比較できる“カスケードグラフ”(付図 $7$ 参照)を表示して、水温逆転なども系統的なものかどうか判断しながら検査する方針をとっている。

これまで述べた検査は計算機を利用したシステムによって実施されている。その概要は付録を参照されたい。

## 10. 論 議

上述の水温の検査方法では、観測された水温データと、標本データとの比較によるもので、同時期に観測された近傍の観測値との比較は“カスケードグラフ”のみである。異常なデータは実際に等値線を描いてみると、その場所で異常な等値線として明瞭に検出できることがある。よって、検査済みのデータで等値線を描き、異常な場所周辺のデータを一つ一つ再検討することが望ましい。多くの場合、孤立した何重もの同心円状の等値線が現れるが、Pazan et al. (1987)は、これを“bull eye”と呼び、この原因となるデータを品質管理の段階で削除している。現在のところ、大洋規模の水平的な等値線を描

くのは、国外・国内の様々な機関からのデータが揃って以後に行うので、当所の観測網から得られるデータのみで等値線によるデータの検査は業務的には実施していない。

当所のような、多数の船を用いて観測網を展開する機関では、データの品質管理は現場への緊密なフィードバックが重要であろう。実際、品質管理で疑わしいデータは観測者に問い合わせ、修正できる場合も少なくない。基本的には、品質管理は観測の現場で行なうのが最も効果的である。なぜなら測器の精度やその作動状態は観測者が最も良く知り得る立場にあるからである。XBT 観測を例に上げれば、ワイヤーのもつれによる延び、船体との接触による切断、絶縁被覆部の漏電などで異常な水温プロファイルが得られるが、現場で判断できれば、観測をやり直して、正しいデータを取り戻すことができる。専門家でなくても現場でプロファイルを診断して、正常に作動しているかどうか判断できる実用的なマニュアルも良いものが出ている (Bailey et al., 1994)。

観測終了後、データだけでなく、同じ海域・時期の標本データの平均的な鉛直プロファイルも表示するようなシステムを作成すれば、上記の品質管理と同様な検査が可能であるし、現在の海況の評価も可能になる。ただし、異常な海況のために正しいデータも異常と判断されることもあるので、疑わしいデータの場合はその場で再度観測したり、測器の状態を確かめるなどの対策をとり、疑わしいデータ自体は排除しない方がよい。

また、人為誤差を少なくするには、位置・時刻などは航海測器から水深/水温データにできるだけ自動的に入力することが望ましい。さらに、船上でデータファイルを作成するソフトにデータの検査機能を付加すれば、観測の現場でデータの品質をチェックできる。観測網を展開している機関では、品質管理はデータ取得の最前線でチェックするシステムの開発が必要であろう。

## 11. ま と め

大洋規模の海洋変動（空間スケール数100km，時間スケール1～2ヶ月）の把握を目的として、観測網を展開する機関における水深/水温データの品質管理方法について考察した。主要な論点は以下の通りである。

- 1) データの品質管理は、データを使用する研究の目的によって決まる。
- 2) 大洋規模の海洋変動をモニターする観測網にとって、現状の観測点の密度から得られる水温分布の推定精度に比べ、水温測器精度はいずれの測器でもあまり問題にならない。ただし、XBT は深度補正が必要である。
- 3) 人為誤差と、測器の不調などに対応する品質管理が必要で、このためにはデータ形式、多重登録、位置/時間の検査が必要である。
- 4) 水深/水温の検査には、これと比較する標本データが必要で、品質管理には、この標本データの統計的性質に合致した取捨選択基準を採用する。
- 5) 太平洋における緯経度5度の区画の季節毎の水温ヒストグラムは、概ね単峰型であるが歪が大きく正規分布をはみせないものが多数（85-90%）を占める。
- 6) 取捨の基準としては、 $\pm 3\sigma$  や、ヒストグラムの適当なパーセント点（0.5%、99.5%など）を決めて、それらの範囲外のを抽出し、人間が再検討するのが妥当であろう。
- 7) さらに、品質管理済みのデータで実際に分布図を描いて異常な水温分布パターンを検出することが望ましい。
- 8) 将来的な品質管理の方向としては、観測の現場において自動的に行えるシステムの作成を目指すべ



きであろう。

## 謝 辞

まぐろ資源調査時の海洋観測資料の管理は、遠洋水産研究所海洋・南大洋部低緯度域海洋研究室が業務として実施している。本研究もこの業務の一環として行なわれたものである。システムの設計段階では、当研究室の渡邊朝生博士から助言を得、実際の業務に関わっている山岸久子氏からも具体的な問題点の指摘を得た。システム開発にあたっては、遠洋水産研究所が基本設計を行ない、東芝 KN システム(株)がコーディングを行なった。

## 付録. 品質管理プログラムの概要

本文で述べたように、観測資料には様々な種類の誤差が含まれている。扱う水深／水温データが年間10,000件近くになると、従来のような人手による個々の資料のチェックは、既に限界に達している。そこで、この作業の自動化を行なうことにした。ただし、ある程度は人間の判断も必要なので、機械と人間が同時に効率よく作業できるようなシステムの開発が必要となる。よって、機械的な判断は機械に任せ人間の判断が必要な時は機械の判断結果をコンピューター画面上の図表によって表現し、人間が最終的な判断を行なうようなシステムが望まれる。遠洋水産研究所では、このようなプログラムを設計・開発して業務に使用している。このシステムの概要を示すことで、本文で述べた考え方がより具体的になると思われる。以下にこのシステムを紹介する。

### 1. プログラムの作動環境

コンピューターシステム

- |                  |                                |
|------------------|--------------------------------|
| 1) コンピューター:      | ヒューレットパッカード社ワークステーション HP715/50 |
| 2) オペレーティングシステム: | HP-UNIX 9.0                    |
| 3) プログラム語:       | FORTRAN77 および C                |
| 4) プログラム環境:      | Xwindow (X11), OSF Motif       |

### 2. プログラムの内容

海洋観測データ(水産庁統一フォーマット, POD形式)の品質管理を自動的に実施する。データはアスキー形式で、船名、位置、観測時刻、気象観測値などを含む3行のヘッダーと、それに続き観測層毎に水深、水温、塩分値が1行をなす可変長のデータである。これを1データとすれば、ファイル中でデータがどのような順番で並んでいても構わない。

プログラムは以下の3つのチェックがあり、通常はこれらを順番に実施して行かなければならない。

フェーズ1:

データのレコード長の検査と、各データ項目について、あり得べき範囲を越えたデータを検出する。数値に関してはほとんど最大最小で検査され(緯度は0~90, 水温は-2~32など)、船名コードや測器コードなどは対照表のファイルを参照してこれに無いものは誤りとみなす。誤りのあるデータはエラーリスト(付表1)として印刷出力され、画面上ではデータの修正を行なうためのエディターウィンドウが開き、誤りのあるデータの行の後にエラーメッセージが付加される(付図1)。このエディターでエラーを編集していく。

エラー修正後、再度このチェックを行ない、エラーが無くなるまで繰り返す。付加されたエラーメッセージは修正が終了した時点で削除されてデータファイルが作成される。この時、デー

Appendix Table 1. Error list of phase 1 check.

\*\*\* OBSERVATION DATA QUALITY CONTROL SYSTEM \*\*\*  
LIMIT & TABEL CHECK ERROR REPORT (CHECK\_1)

DATE : 1994.12.12  
FILENAME : ./sample.cur

ERROR CODE	ERROR LINE NO.	RECORD NO.	VESSEL CODE	ST. NO.	DATE	ERROR MESSAGE
0001	0004	0003	990	T 1	92.09.21	INVALID RECORD LENGTH
0001	0073	0072	990	T 4	92.09.22	INVALID RECORD LENGTH
1004	1386	1385	990	F46	92.11.18	FIND INVALID VALUE AT 27 COLUMN
3001	1388	1387	990	F46	92.11.18	BLANK DATA AT 18 COLUMN
1004	1410	1409	990	F47	92.11.19	FIND INVALID VALUE AT 27 COLUMN
1004	1431	1430	967	T 1	92.09.24	FIND INVALID VALUE AT 27 COLUMN
1004	1450	1449	967	T 2	92.09.24	FIND INVALID VALUE AT 27 COLUMN
1004	1469	1468	967	T 3	92.09.24	FIND INVALID VALUE AT 27 COLUMN
3001	1472	1471	967	T 4	92.09.25	BLANK DATA AT 11 COLUMN
3005	1474	1473	967	T 4	92.09.25	FIND INVALID CODE DATA AT 14 COLUMN
2001	1492	1491	967	T 5	92.09.25	BLANK DATA AT 29 COLUMN
5005	1574	1573	967	T 9	92.09.26	FIND INVALID CODE DATA AT 17 COLUMN
3005	1602	1601	967	T 11	92.09.26	FIND INVALID CODE DATA AT 14 COLUMN
5005	1765	1764	967	T20	92.09.29	FIND INVALID CODE DATA AT 17 COLUMN
3001	2329	2328	967	F21	92.11.02	BLANK DATA AT 18 COLUMN

TOTAL NUMBER OF ERROR : 15

The screenshot shows an editor window with a list of data points and an error message. The data points are numbered 1 through 23, each with a record ID, a vessel code, and a date. The error message is on line 4, indicating an invalid record length. The window title is './sample.cur' and it shows 3943 lines and 127596 characters.

1	182230111430905990	T192092118	0182835248N163507E09
2	18223022	ENE1	E1 222 199 C ENE 3 210192
3	18223033DBT	+	MIZUNO
4	*** ERROR ***	INVALID RECORD LENGTH	
5	18223045	0	2300034360
6	18223055	10	22888
7	18223065	20	22870
8	18223075	30	22850
9	18223085	50	18579
10	18223095	75	16530
11	18223105	100	14909
12	18223115	125	13720
13	18223125	150	12829
14	18223135	175	11829
15	18223145	200	11199
16	18223155	250	10170
17	18223165	300	08390
18	18223175	400	05638
19	18223185	500	05250
20	18223195	600	04360
21	18223205	700	04050
22	18223215	800	03730
23	18223225	900	03420

"./sample.cur" 3943 行, 127596 文字

Appendix Figure 1. Editor window for phase 1 check. Comments are appended to erroneous lines.

**Appendix Table 2.** Summary report of phase 1 check, specifying the observations by ship and observation mode (Transect/Fishig ground).

## \*\*\* OBSERVATION DATA QUALITY CONTROL SYSTEM \*\*\*

## SUMMARY REPORT

DATE : 1994.12.12

FILENAME : ./sample.cur

OFFICE CODE	VESSEL CODE	ATT.	CRUISE No.	OBSERVATION TERM	TOTAL St. No.	(CLASSIFIED STATION NUMBER BY INSTRUMENT)							
						TEMP.	TEMP.	TEMP.	SAL.	SAL.	CURRENT	CURRENT	
2432	967 ○○丸	T	02	92.09.24 - 92.09.29	22	XBT	22			BTL	2	DPL	22
2432	967 ○○丸	F	02	92.09.30 - 92.11.23	35	DBT	31	RTM	4	BTL	4	DFT	35
2530	969 ××丸	T	02	92.09.23 - 92.09.30	28	CTD	28			CTD	28		
2530	969 ××丸	F	02	92.10.05 - 92.11.23	31	CTD	31			CTD	31	DFT	26
1430	990 △△丸	T	01	92.09.21 - 92.09.26	20	DBT	19						
1430	990 △△丸	F	01	92.09.21 - 92.11.19	47	DBT	42	RTM	5	BTL	5	DFT	47

TOTAL NUMBER OF STATION : 183

TOTAL NUMBER OF RECODE : 3928

TOTAL NUMBER OF ERROR : 155

タは船別・時間順にソートされている。この段階で、観測状況を船別・航海別に測器の種類と各々の測器毎の測点数などが一覧表（付表2）として報告される。

## フェーズ2：

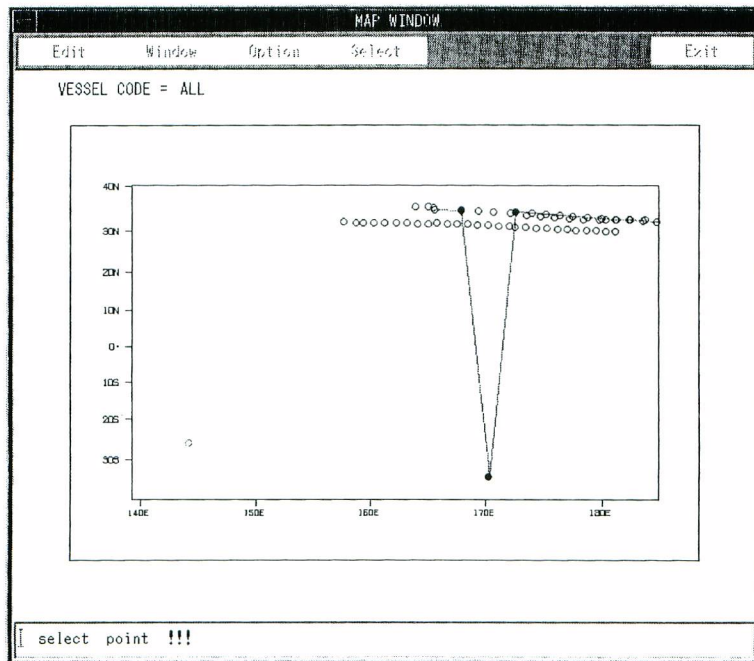
船名コード順，時間順にデータを並べ，その位置・時間に矛盾がないか検査する。ウィンドウ上に各船毎に測点を順に結んだ航跡図が作成され，測点間の距離を移動に要した時間で割って船速が計算され，これが“基準速度”を越えた点の一覧表（エラーリスト）が印刷出力される（付表3）。一方，ウィンドウ上ではこの異常を検出した点は赤色（正常なものは黒色）で示される（付図2）。この点をマウスで指定すると，修正ウィンドウが開き，該当する点の位置，日時，日付，船名コードが数値で示され（付図3），これらを自由に修正できる。このチェックをエラーリストが出なくなるまで修正を繰り返す。

Appendix Table 3. Error list of phase 2 check.

\*\*\*\*OBSERVATIONDATA QUALITY CONTROL SYSTEM\*\*\*\*  
 CHECK INCONSISTENCY OF OBSERVATION POINT & TIME ERROR REPORT (CHECK 2)  
 DATE : 94.12.13

FILENAME: ../sample.cur

ERROR CODE	SERIAL NO.	VESSEL CODE	VESSEL NAME	INSTRUMENTS		DATE	TIME	STATION		ERROR MESSAGE
				TEMP	SALT			LATITUDE	LONGITUDE	
6002	144	969	F-20	CTD	CTD	92.11.07	09:38	26:33.0 N	151:33.5 W	IRREGULAR OBSERVATION POINT TO THE NEXT POINT
6001	145	969	F-21	CTD	CTD	92.11.08	09:59	25:05.3 S	152:37.0 W	IRREGULAR OBSERVATION POINT TO THE PREVIOUS POINT
6002										IRREGULAR OBSERVATION POINT TO THE NEXT POINT
6001	146	969	F-22	CTD	CTD	92.11.09	09:55	24:37.9 N	152:41.6 W	IRREGULAR OBSERVATION POINT TO THE PREVIOUS POINT
6002	151	969	F-27	CTD	CTD	92.11.16	10:15	31:40.0 N	159:14.1 W	IRREGULAR OBSERVATION POINT TO THE NEXT POINT
6001	152	969	F-28	CTD	CTD	92.11.17	09:45	31:01.1 N	161:08.1 E	IRREGULAR OBSERVATION POINT TO THE PREVIOUS POINT
6002										IRREGULAR OBSERVATION POINT TO THE NEXT POINT
6001	153	969	F-29	CTD	CTD	92.11.18	09:41	33:07.6 N	160:45.2 W	IRREGULAR OBSERVATION POINT TO THE PREVIOUS POINT
6002	182	990	T1	DBT		92.09.21	18:00	35:24.8 N	163:50.7 E	IRREGULAR OBSERVATION POINT TO THE NEXT POINT
6001	183	990	T2	DBT		92.09.22	00:00	35:17.5 S	164:57.3 E	IRREGULAR OBSERVATION POINT TO THE PREVIOUS POINT
6002										IRREGULAR OBSERVATION POINT TO THE NEXT POINT
6001	184	990	T3	DBT		92.09.22	06:00	35:02.1 N	16:00.0 E	IRREGULAR OBSERVATION POINT TO THE PREVIOUS POINT
6002										IRREGULAR OBSERVATION POINT TO THE NEXT POINT
6001	185	990	T4	DBT		92.09.22	11:58	34:45.1 N	167:49.8 E	IRREGULAR OBSERVATION POINT TO THE PREVIOUS POINT



Appendix Figure 2. Observation station window for phase 2 check. A Track line and observation stations for the ship checked are shown. Solid circles denote erroneous position/time data station with neighbour stations.

MODIFY WINDOW		
OK	Delete	Cancel
VESSEL CODE	967	〇〇丸
DATE YEAR/MONTH/DAY	920924	
TIME : START TIME	1600	
END TIME	1603	
POSITION LATITUDE	-34.270	
LONGITUDE	167.515	

**Appendix Figure 3.** An window to modify erroneous position/time data. Pointing any station in the window in the previous figure opens this window.

フェーズ 3 :

水温・塩分に関しては、統計値と比較して余り大きな差があるものでは、問題があるものと判断して、それらのデータがエラーリスト（付表 4）として報告される。統計値緯経度 5 度区画の小領域毎に、所定深度の季節別平均値と標準偏差を計算したものを使用している (Mizuno and Yukinawa, 1991)。ある観測値と対応する平均値との差の大きさの判断基準は通常標準偏差の 3 倍を使用しているが、この倍率は自由に指定できる。

このチェックでは、付図 2 と同様なウィンドウ上に観測点の地理的な分布が表示され、問題があるとみなされた観測点は赤色で示される。この点をマウスで指定すると、観測した鉛直プロフィールが平均値のプロフィールとともに示される (付図 4)。平均値のプロフィールには標準偏差と判断基準の倍率に対応する幅を持つ水平な線分が各所定層に描かれている。この線分の実線部は  $\pm\sigma$  ( $\sigma$  は標準偏差)、破線部は  $\pm 3\sigma$  の範囲である。この破線の範囲からはずれたデータはその深度で赤色の点で示される。この図により、平均値からの違いの度合いとプロフィール自体から、最終的に人間が修正・削除の判断を下す。

修正を要すると判断した時は、ウィンドウ上の鉛直プロフィールをマウスでクリックすると、このプロフィールに対応する水深、水温、塩分の数表が表示される (付図 5)。修正すべきデータをこの表から指示すると修正用のウィンドウが表示され (付図 6)、ここで修正すべきデータの編集 (変更、削除など) を行なう。

以上の他、一連の測点毎に水温または塩分の鉛直プロフィールのグラフを並べた“カスケードグラフ” (付図 7) も表示させることができる。これは、ある測点のプロフィールと、その前後の測点のプロフィールが同時に表示されるので、その点のみが異常なプロフィールになっているかどうかが目瞭然である。これは、上記の平均値との比較では見出せなかった異常なプロフィールも検出できる。ただし、この判断自体は人間が行なわなくてはならない。

Appendix Table 4. Error list of phase 3 check.

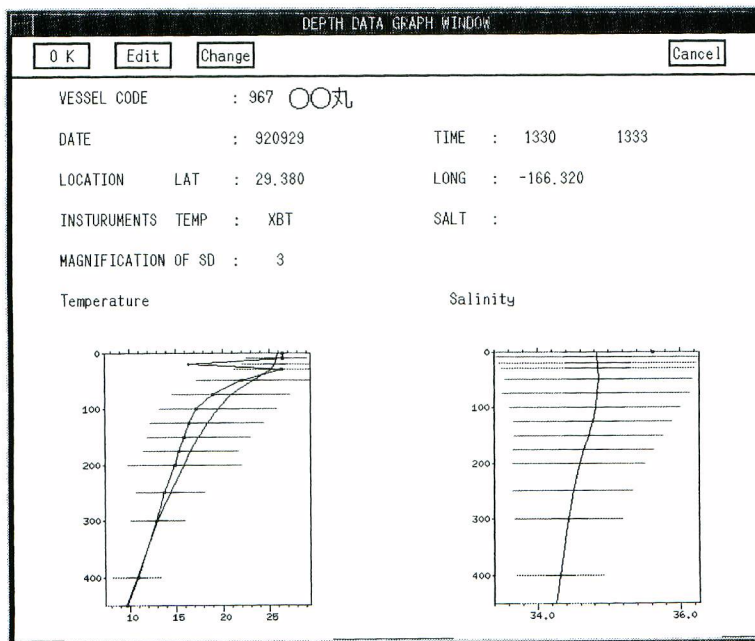
\*\*\*\*OBSERVATION DATA QUALITY CONTROL SYSTEM\*\*\*\*

COMPARE OBSERVATION DATA WITH CLIMATOLOGICAL DATA ERROR REPORT (CHECK 3)

DATE : 94.12.14

FILENAME: ./sample.cur

ERROR CODE	SERIAL NO.	VESSEL CODE	ST. REC. NO. NO.	DATE	TIME	LATI TUDE	LONGI TUDE	INSTRUMENT	MIN/ MAX MEAN S D	MIN/ MAX MEAN S D	SALT								
											TEMP.	SAL.	DEPTH	ANGLE	TEMP.	SALT.	MAX	MEAN	S D
7002	142	967	T22	6	92.09.29	13:30	29:38.0 N	166:32.0 W	XBT	LOW	20	16.399	22.050	25.620	1.190				
7002	144	967	F1	19	92.09.30	06:10	29:31.3 N	165:14.4 W	DBT	LOW	600	5.960	6.600	6.600	0.000				
7001	147	967	F4	15	92.10.08	06:55	25:40.3 N	150:20.1 W	DBT	HIGH	250	16.819	16.580	13.160	1.140				
7001	158	967	F14	21	92.10.26	06:25	27:02.4 N	148:12.3 W	DBT	HIGH	800	14.380	4.850	3.950	0.300				
7001	161	967	F17	19	92.10.29	06:25	28:13.6 N	146:54.0 W	DBT	HIGH	600	5.980	5.960	5.450	0.170				
7002	168	967	F23	19	92.11.04	06:15	26:43.3 N	145:39.6 W	DBT	LOW	600	4.789	4.940	5.450	0.170				
7011	136	969	F-12	4	92.10.25	09:37	21:54.1 N	149:52.3 W	CTD	HIGH	0					37.740	35.890	34.810	0.360
7002	136	969	F-12	23	92.10.25	09:37	21:54.1 N	149:52.3 W	CTD	LOW	1000	3.930	3.990	4.170	0.060				
7002	137	969	F-13	23	92.10.26	09:47	21:46.6 N	149:23.2 W	CTD	LOW	1000	3.980	3.990	4.170	0.060				
7002	138	969	F-14	23	92.10.27	09:46	21:19.3 N	149:39.0 W	CTD	LOW	1000	3.970	3.990	4.170	0.060				
7002	139	969	F-15	23	92.10.28	09:39	22:08.7 N	149:50.5 W	CTD	LOW	1000	3.970	3.990	4.170	0.060				
7002	150	969	F-26	20	92.11.14	09:53	26:39.5 N	160:12.5 W	CTD	LOW	700	4.570	4.630	5.305	0.225				
7001	229	990	F28	04	92.01.29	11:30	28:12.8 N	148:53.1 W	RTM	HIGH	0	24.298	23.850	21.060	0.930				
7002	203	990	F2	19	92.09.29	12:01	29:19.4 N	165:51.0 W	DBT	LOW	600	6.360	6.600	6.600	0.000				
7001	204	990	F3	19	92.09.30	12:10	28:36.6 N	165:32.7 W	DBT	HIGH	600	6.769	6.600	6.600	0.000				



Appendix Figure 4. An window showing observed erroneous temperature/salinity vertical profiles found in phase 3 check. In both profiles, climatological profiles are shown. Solid and broken horizontal bars at standard depths denote the range of 1 and 3 standard deviation range respectively. In terms of salinity, only surface data was obtained.

DEPTH DATA WINDOW

OK Cancel

VESSEL CODE : 967 ○○丸

DATE : 920929

TIME : 1330 - 1333

LOCATION : 29.380 -166.32

DEPTH ANGLE	TEMPERATURE		SALT
	Min	Max	
0	26.500	26.500	35.631
10	26.500	26.500	
20	22.050	16.399 29.190	TEMP ERROR
30	26.298	26.298	
50	22.000	22.000	
75	18.899	18.899	
100	17.200	17.200	
125	16.399	16.399	
150	15.899	15.899	
175	15.300	15.300	

Appendix Figure 5. An window showing the same data as shown in the the previous figure for specifying the data to edit.

MODIFY WINDOW

OK Delete Cancel

VESSEL CODE 967 ○○丸

DATE 920929

TIME 1330 - 1333

LOCATION LAT 29.380 LONG -166.320

INSTURUMENTS TEMP XBT SALT

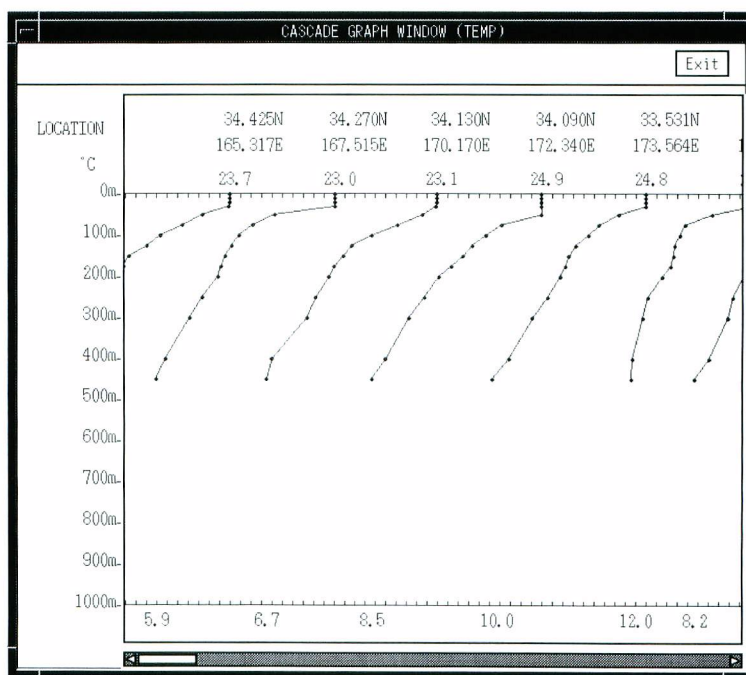
DEPTH/WIRELENGTH 20

ANGLE

TEMPERATURE 16.399

SALT

Appendix Figure 6. An window to edit the erroneous depth/temperature/salinity data.



Appendix Figure 7. Consecutive vertical profile so called “cascade graph” for temperature. It makes easier to find erroneous data.

## 参 考 文 献

- BAYLEY, R., A. GRONELL, H. PHILLIPS, E. TANNER and G. MEYERS 1994: Quality control cookbook for XBT data (version 1.1). *CSIRO Marine Laboratories Report 221*, Hobart, 84pp.
- BOYER, T. and S. LEVITUS 1994: Quality control and processing of historical temperature, salinity, and oxygen data. *NOAA Technical Report NESDIS 81*, Washington D. C., 64pp.
- 大日本水産会 1990: 漁船測位システム調査検討事業報告書 (平成元年度). 104pp.
- FLIERL, G. and A.R. ROBINSON 1977: XBT measurements of thermal gradient in the MODE eddy. *J. Phys. Oceanogr.*, 7, 300-302.
- GANDIN, L.S. 1963: Objective analysis of meteorological field. *Gidrometeorologicheskoe Izdatel'stvo*, Leningrad, U.S.S.R., 286pp.
- GLANTZ, M.H. 1992: *Climate Variability, Climate Change, and Fisheries*. Cambridge University Press, Cambridge, 455pp.
- HANAWA, K. and Y. YORITAKA 1987: Detection of systematic errors in XBT data and their correction. *J. Oceanogr. Soc. Japan*, 43, 68-76.
- HEINMILLER, R.H., C.C. EBBESMEYER, B.A. TAFT, D.B. OLSON, and O.P. NISKIN 1983: Systematic errors in expendable bathythermograph (XBT) profiles. *Deep-Sea Res.*, 30, 1185-1196.



- HALLOCK, Z.R. and W.J. TEAGUE 1992: The fall rate of the T-7 XBT. *J. Atmosph. and Oceanic Tech.*, **9**, 470-483.
- 広辞苑 1976: 新村 出 編 (第二版補訂版) 岩波書店, 東京.
- LANDER, M.A. and M.L. MORRISSEY 1987: Unexpected duplicate ship reports in the comprehensive Ocean-Atmosphere Data Set (COADS). *Tropical ocean-atmospher newsletter*, (38), 13-14.
- MEYERS, G., H. PHILLIPS, N. SMITH and J. SPRINTALL 1991: Space and time scales for optimal interpolation of temperature—Tropical Pacific Ocean. *Prog. Oceanogr*, **28**, 189-218.
- MIZUNO, K. and W.B. WHITE 1990: Volunteer observing ship XBT program (Australia, France, Japan, and U.S.A.): Redistribution of water mass in the western tropical Pacific associated with the ENSO cycle from 1967-1986. Proceedings of international TOGA Scientific conference.
- MIZUNO, K. and M. YUKINAWA 1991: Seasonal vertical temperature profile in the tropical Pacific Ocean. *Bull. NRIFS S-ser.*, (19), 178pp.
- 水野恵介 1994: 海洋観測に関するアンケート調査結果. 平成5年度マグロ漁業研究協議会報告書, 水産庁・遠洋水産研究所, 218pp.
- 奥野忠一 1986: 統計学辞典. 裳華房, 東京, 856pp.
- PAZAN, S. and W.B. WHITE 1993: JEDA center annual report on global sub-surface thermal structure-1989. SIO Ref. No. 93-38. 55pp.
- ROEMMICH, D. and C. WUNSCH 1984: Apparent changes in the climate state of the deep North Atlantic Ocean. *Nature*, (307), 447-450.
- STURGES, H.A. 1926: The choice of a class interval. *J. Amer. Stat. Assoc.*, **21**, 65-66.
- 鈴木栄一 1968: 気象統計学. 地人書館, 東京, 314pp.
- 行縄茂理, 水野恵介, 宮部尚純, 鈴木治郎, 小井土隆, 清田雅史, 西川康夫, 藁科侑生 1988: まぐろ類の再生産および分布に与えるエル・ニーニョの影響. 昭和62年重点基礎研究成果報告書, 1-13.
- WHITE, W.B. and R. BERNSTEIN 1979: Design of an oceanographic network in the midlatitude north pacific. *J. Phys. Oceanogr.*, **9**, 592-606.
- WMO 1994: 私信
- WYRTKI, K. 1975: El Nino—The dynamic response of the equatorial Pacific ocean to atmospheric forcing. *J. Phys. Oceanogr.*, **4**, 572-584.