An Oceanic Subsurface Thermal Analysis Scheme with Objective Quality Control

Neville R. SMITH

Bureau of Meteorology Research Center
Melbourne, Vic. 3001 - Australia.

1. Introduction

In view of the strong impact of interannual climate variability in the Australian region in the form of droughts and other extreme weather events, and of the strategic proximity of Australia to climatically sensitive regions of the tropical oceans, the Bureau of Meteorology Research Centre and CSIRO Division of Oceanography have jointly initiated a project to provide an oceanic analysis system for the tropical Indian and Pacific Oceans. The primary function of the project is to provide timely local analyses with scope for extra attention to the Indonesian Archipelago and Western Pacific regions. This paper will concentrate on the objective quality control procedures that have been implemented with the scheme, using the 1986-87 period to illustrate their operational function. Details of the methodology and implementation are given elsewhere [Blomley et al. 1989]. We hope to provide to tropical ocean analysis the variety of approach that has proved invaluable in the meteorological community.

2. The analysis scheme

Statistical (or optimal) interpolation (SI) is used to analyze data from an irregularly spaced observation network onto a regular grid. Such methods have previously been used to analyze short-term climatic variability in the Pacific Ocean [e.g. White et al. 1985]. The SI scheme used here is a univariate adaption of the ECMWF scheme [Lorenc 1981], incorporating elaborate techniques for data checking and validation, and introduces novel methods for data selection and analysis sector merging. Analyses are provided for sea surface temperature, mixed-layer depth, depth-averaged temperature to 150m and 400m, depth of the 15°, 20° and 25° isotherms, and at 15 standard NODC levels to 500m.

SI utilises prior knowledge of the statistical structure of first-guess and observation errors (deviations from the true field) to transport information from the observation points to some regular grid. Let O, F and A be the observed, first-guess (e.g. a model forecast) and analyzed fields, and E^o and E^f be the estimated errors in the observed and guessed fields. Now define normalised observation and analyzed fields,

$$q_i = (O_i - G_i)/E^f \quad (1)$$

$$r_k = (A_k - G_k)/E^f \quad (2)$$

for i=1,N observations and k=1,K grid points. Each observation deviation q_i is weighted according to its

worth as an estimator for the true deviation r_k at grid point k,

$$r_k = \underline{W}_{ik} q_i \tag{3}$$

where $\underline{W}_{ik}$ is an N x K matrix. The weight matrix is determined by the statistical structure through the relationship

$$\underline{W}_{ik} = \underline{M}^{-1}_{ij} h_{jk} , \tag{4}$$

where $\underline{M}_{ij}$ is the correlation matrix and $\underline{M}^{-1}$ is its inverse. h_{jk} is the correlation vector for observations i at grid point k (refer to Blomley et al. 1989 for further detail). The analysis error is given by

$$(e^a_k)^2 = 1 - h_{ik}\underline{M}^{-1}_{ij} h_{jk} . \tag{5}$$

The data for real-time analyses are drawn from the Melbourne node of the GTS in the form of BATHY and TESAC messages. The data base for research quality analyses is supplemented by data supplied by NODC and the TOGA Subsurface Data Centre (TSDC). Observations are checked for reasonableness and exact- and near-duplicates eliminated from the data base.

3. Quality control and data selection

Fig. 1 is a schematic of an idealised data distribution and first-guess estimate in a one-dimensional domain. One of the unique features of the scheme is that it brings to oceanography the powerful methods of data checking routinely used in meteorology, thus eliminating labour-intensive manual checking of data.

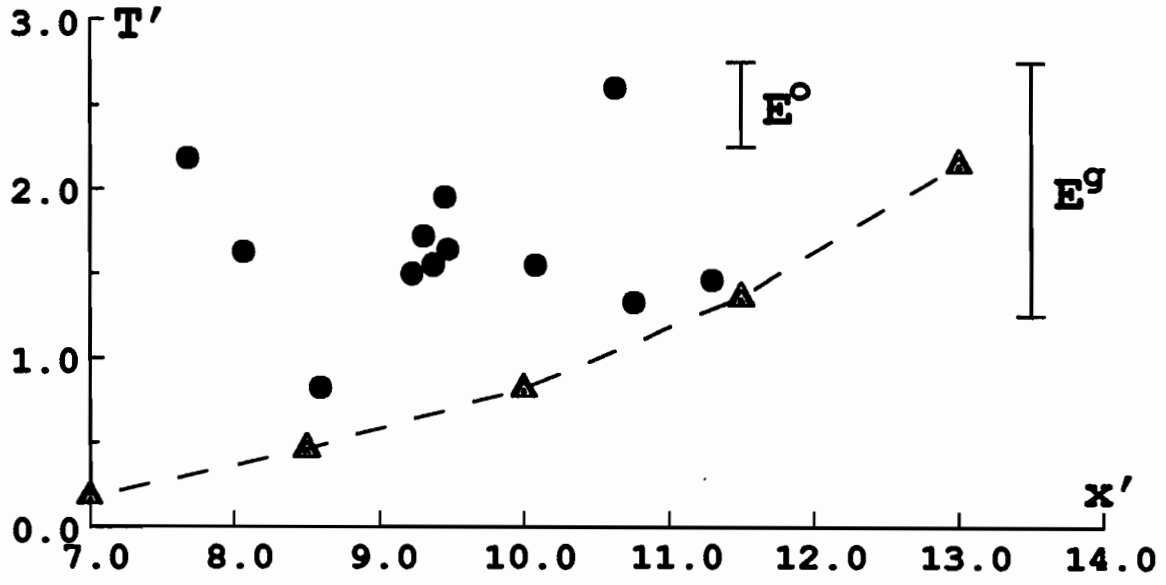


FIG. 1. The distribution of the original observations [●] and the first-guess (dashed curve) for some arbitrary field T'. The decorrelation scale is 1.5 in the arbitrary space units of x'.

Observations are stored internally as normalised deviations from a first-guess field, usually derived from climatology [e.g. Levitus 1984]. In this case the estimated error of the first-guess E^g is simply the standard deviation of the true field. E^o is the standard deviation of the observations about the true field (i.e. the instrumental and geophysical noise). Observations are pre-checked within the analysis using a

3-sigma test against the first-guess (Fig. 2), each suspect observation being flagged and temporarily omitted.

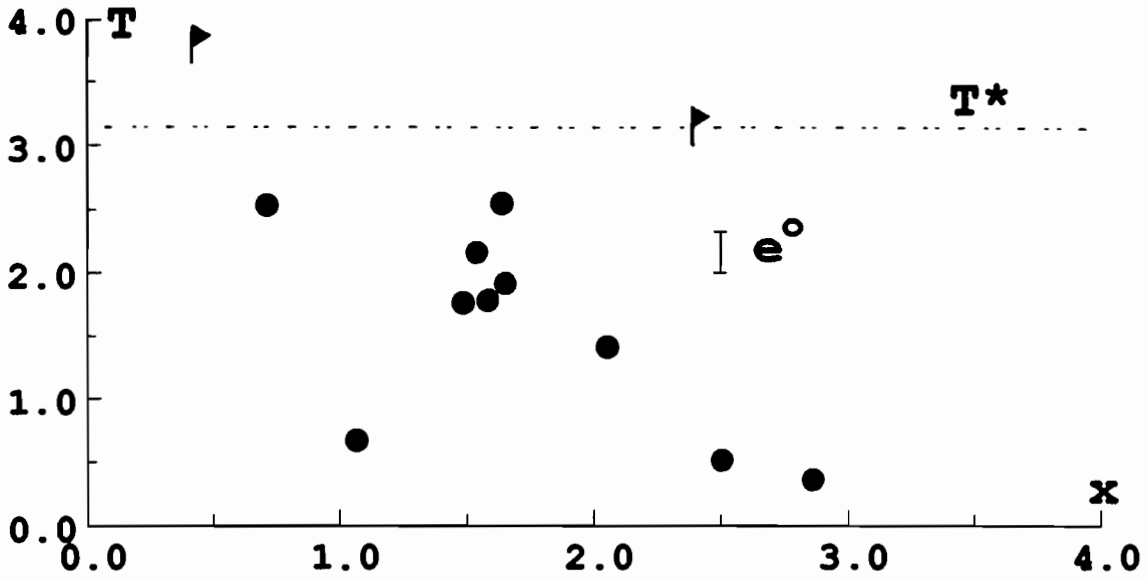


FIG. 2. A map of the observations in the normalised domain. The tolerance T' is a function of the noise-to-signal ratio e' and the rejection criteria R'. Suspect observations are flagged [▶].

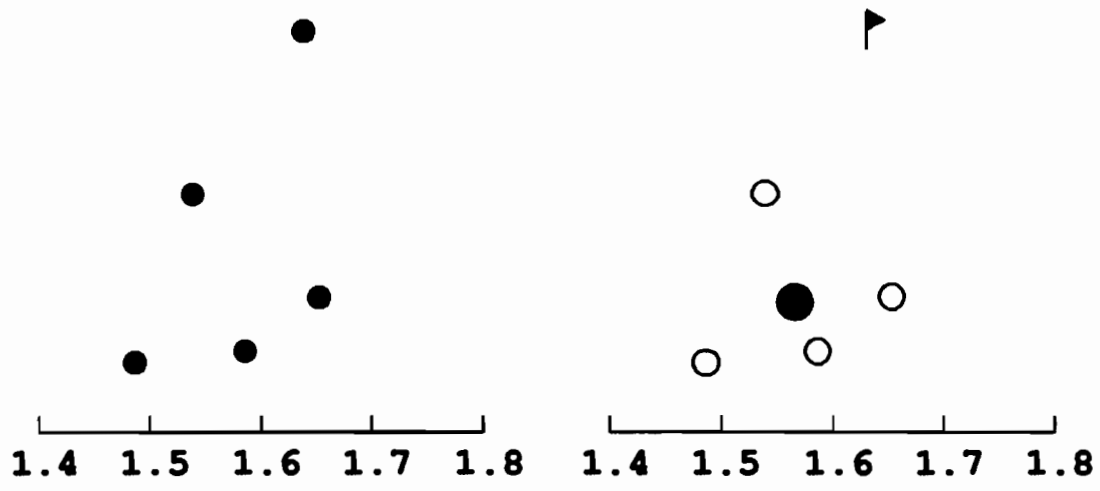


FIG. 3. Schematic of super-observation formation. A group of close observations [●] (left panel) are checked and suspect data are flagged (right panel). A new observation [◉] is formed from the redundant data [○].

Next, close groups of data - here "close" means the data are statistically indistinguishable - are considered for combination into super-observations (Fig 3). Modified SI equations are used to inter-calibrate the individual observations against the expected super-observation value. Redundant data are cancelled and suspect data temporarily flagged. Super-observations are henceforth considered to be of superior accuracy (smaller error) compared to their constituent data. Both the cross-validation and final

analysis are instituted through an arrangement of inter-meshed analysis sectors (Fig. 4). The number of observations will usually exceed the numerical capability for matrix inversion and the domain must be subdivided into manageable analysis sectors. This arrangement ensures all observations are selected at least once. The inter-meshing may be adjusted to eliminate the overlap altogether or to encourage multiple-selection of all observations. The weighting between interpolations from different sectors is decided by the estimated error of the individual analyses.

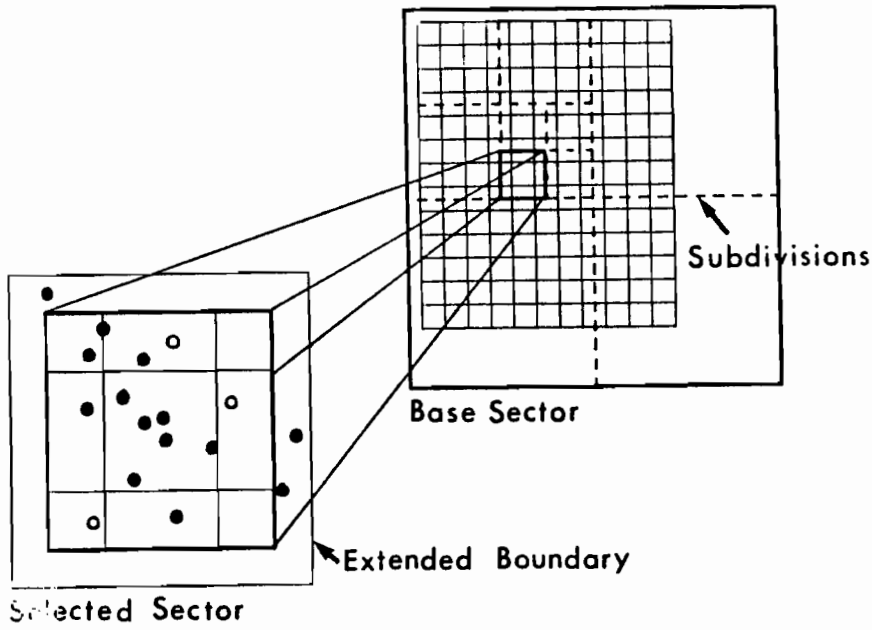


FIG. 4. A schematic of data selection. The analysis grid (upper right) is subdivided until the number of data selected for the sector is optimum (lower left). The inner sector is extended to facilitate sector overlap (outer box).

Remaining single observations and super-observations are cross-validated against the projected grid analysis (Fig. 5) with data incurring two or more flags being permanently deleted. Temporarily flagged data are rechecked and flags made permanent or lifted (Fig. 6). The cross-validation process should not be viewed solely as a mechanism for detecting rogue observations but also as an integral part of analysis "smoothing"; observations which are not consistent with the imposed statistical model may be eliminated.

Finally the analysis is performed over all sectors using the remaining single and super-observations (Fig. 6). In addition to the analyzed increments r_k the analysis error ϵ^a is also estimated. The analysis error is a measure of the relative confidence in the individual estimates and is used to control the merging of interpolations from adjacent sectors as well as to form the analysis error field for the final analysis. The final analysis (Fig. 7) will favour the observations in data rich regions and the error will be small (e.g. near x^*=9.3). In data sparse areas the analysis is near the first-guess and the error approaches the first-guess error (e.g. at x^*=13). If there is only a single observation within a radius of the decorrelation scale, the estimated error will be just less than the observation error.

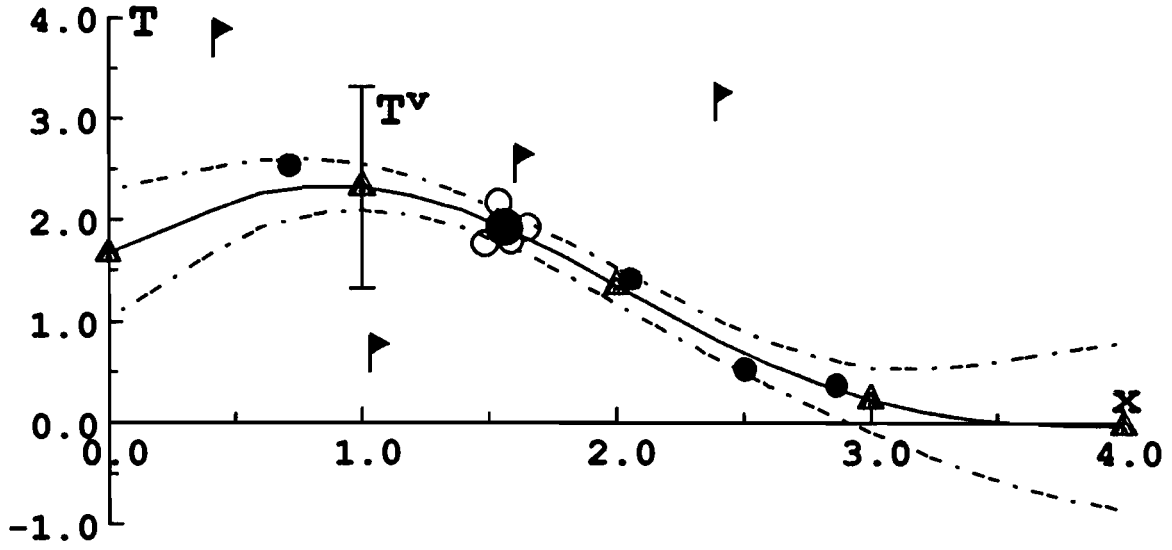


FIG. 5. Observations are checked by cross-validation against an independent analysis [-▲-]. The tolerance T^V is based on the estimated error e^* (dashed) and a rejection criteria R^V. Suspect data (e.g. near x=1) are flagged.

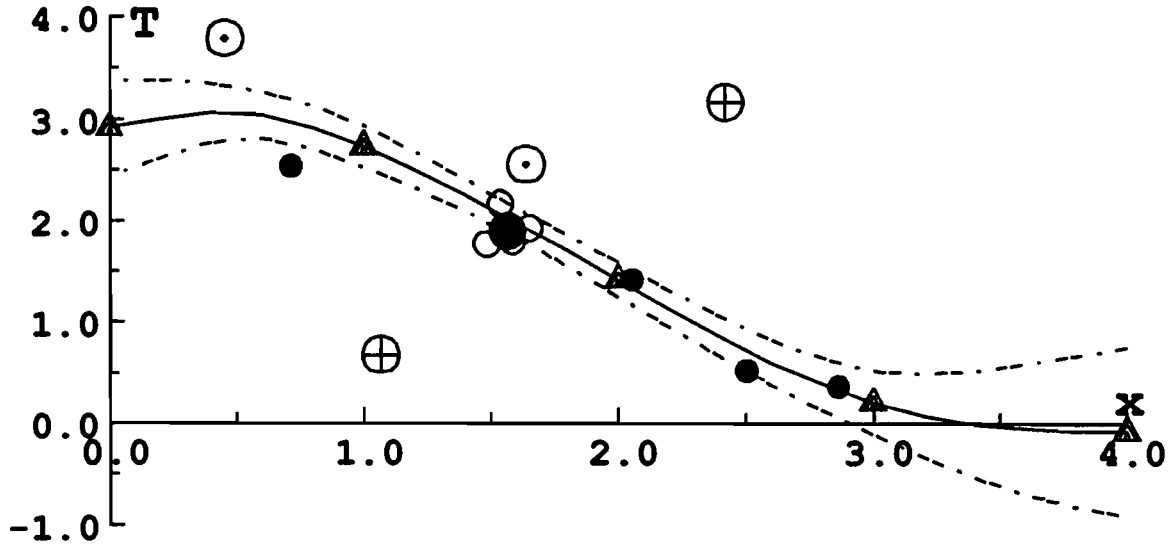


FIG. 6. Flagged data are rechecked against the analysis [▲]. Some data are readmitted [⊙] while failures are deleted [⊕]. The solid and dashed curves are the final normalised analysis and estimated error.

4. Quality control procedures in operation

The operational implementation of quality control procedures is complicated by the non-isotropic, spatially varying nature of the auto-correlation functions for the tropical oceans. Currently the decorrelation scales are 1500 km x 300 km in the equatorial region, varying smoothly to 500 km x 500 km in midlatitudes and 100 km in coastal regions [Meyers et al. 1989]. Particular care must be taken to avoid indeterminate calculations for the analysed field and its associated error. Even then, observations may occasionally be unjustly excluded, usually in regions of dense data coverage or in coastal regions

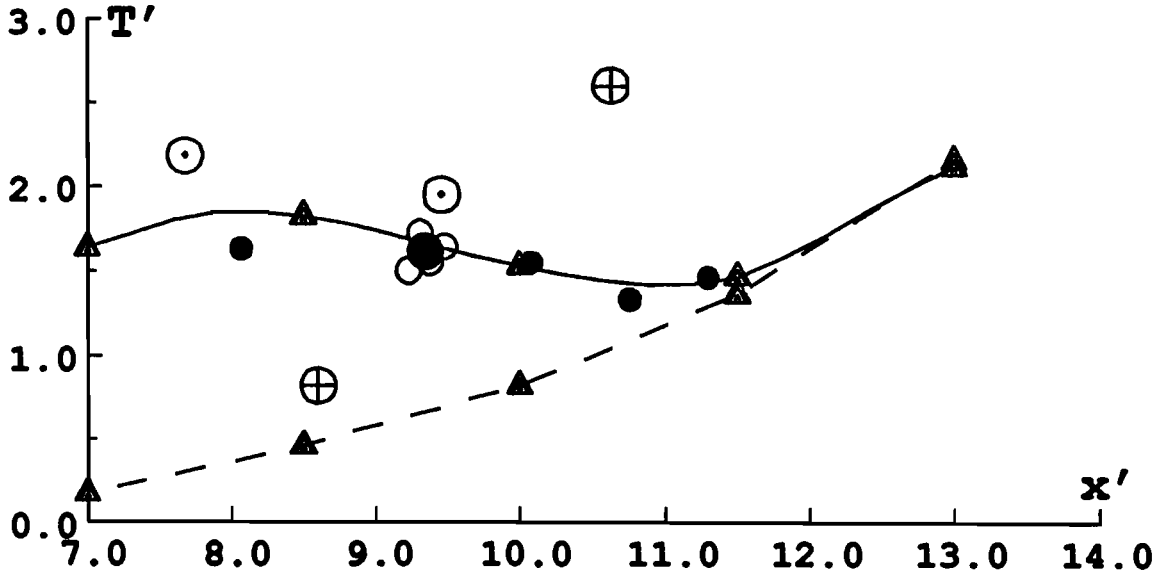


FIG. 7. Final analysis with observations mapped to the original unnormalised domain. [●], [○] and [◉] are single, redundant and super-observations respectively. [⊕] and [⊙] are deleted and validated suspect data.

where the imposed statistical structure varies rapidly. Given sufficient information on which to make an objective assessment, the scheme has proved more than satisfactory as an arbiter of poor measurements, . For the real-time analyses the path from the GTS BATHY/TESAC message data base to the grid point analysis fields is entirely objective, thus avoiding the labour intensive, subjective interventions.

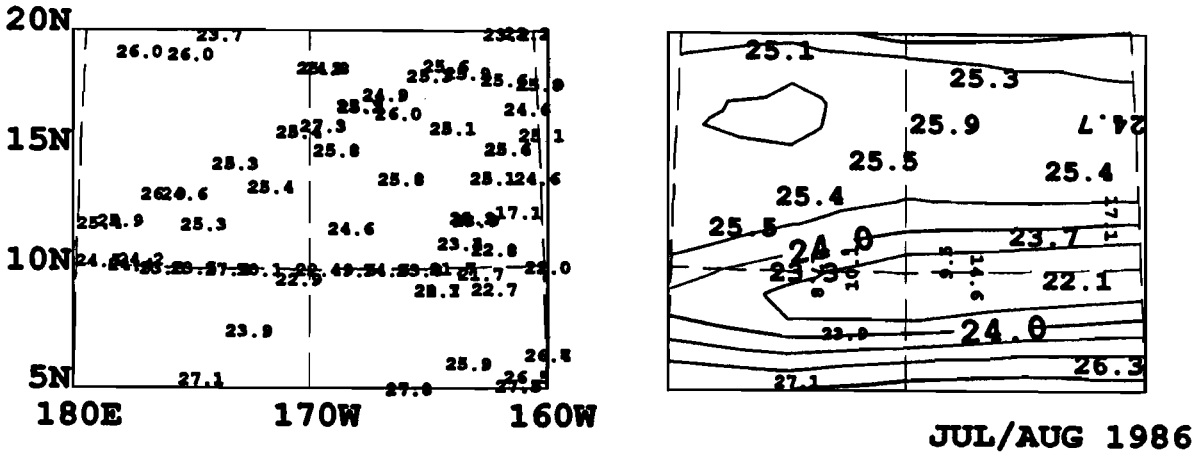


FIG. 8. An expanded view of the T_{150} analysis. The left panel shows the original data while the right panel shows the ultimate distribution together with analysis.

As an illustration of the operational action of the scheme, figure 8 presents details of an analysis over a restricted region of the Pacific for the 150 m depth-averaged temperature (c.f. the schematic examples of Figs 1 through 7). Single and super observations are written horizontally, with super-

observations being shown slightly larger. Deleted observations are rotated through 90°, while suspect (but cleared) data are upside-down. The original data have been duly screened by pre-checking and cross-validation for inconsistent and suspect data. Excessive redundancy has been removed by combining groups of close observations into super-observations. This example clearly demonstrates the degree to which data are payed by the analysis. For ordinary, single observations this is determined by the imposed noise-to-signal ratio. For super-observations, the observational error assigned to the new datum truly reflects the information content - if the constituents were evenly spread in time and space (as determined by the decorrelation scales) then the new datum is considered to be more accurate than its components. Conversely, if the constituents are very close in both space and time (for example, many measurements at a single station) then the newly formed datum will be of similar accuracy to its components. Clearly the efficacy of the quality control procedures depends on the quality of the first-guess field (climatology is by no means ideal), on the imposed statistical structure, and on the many internal parameters of the SI scheme. These dependencies have yet to be examined in detail but experience with the near real-time system, and with analyses of NODC and TSDC data, suggest the present configuration is satisfactory.

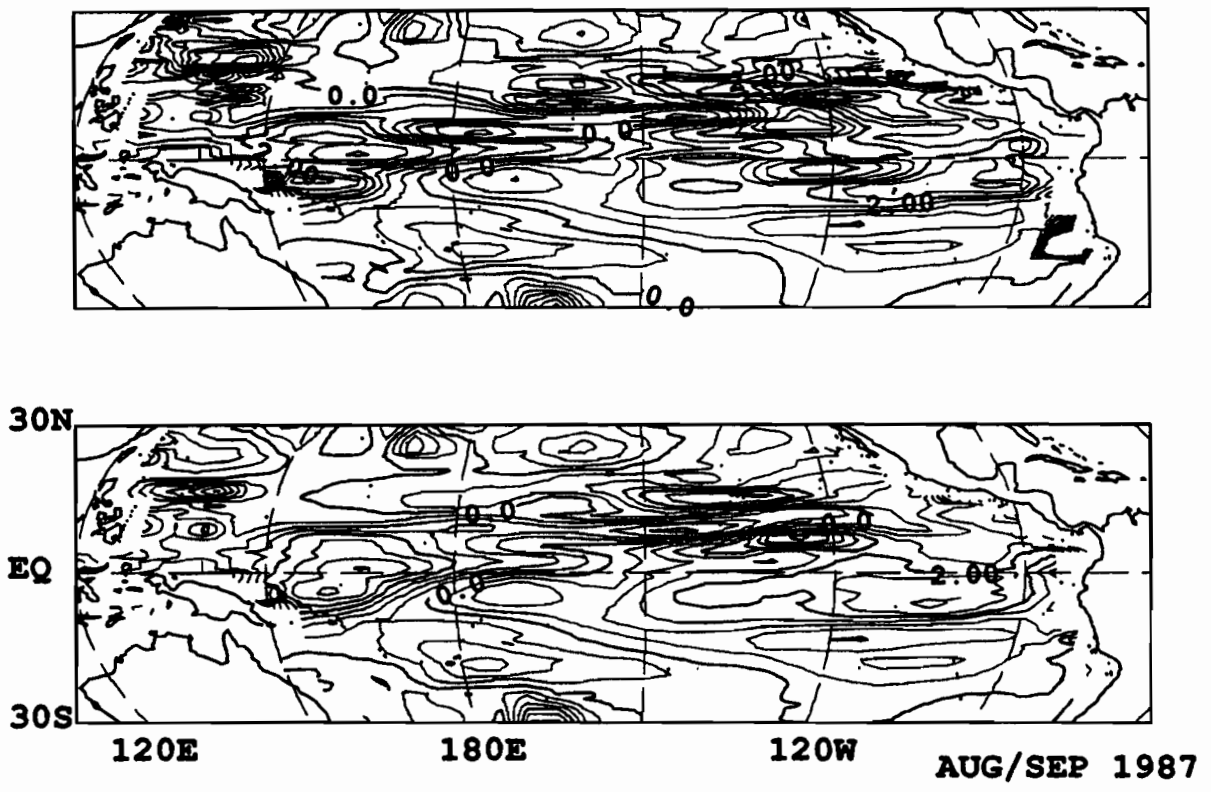


FIG. 9. A comparison between an analysis with no quality control (upper panel) and an analysis with full quality control (lower panel) over the equatorial Pacific. The field is the 150m depth-averaged T_{150} increment.

As a further illustration of objective quality control, figure 9 compares the increments from an analysis with regular controls with an analysis where all controls have been turned off. The data in this case are derived from the BATHY message data base. A marked deterioration in the analysis is evident, particularly east of the dateline in the North Pacific. Some of the changes can be traced to spurious

observations, while others appear to be related to subtleties of the SI scheme. For example, the depression near 18°N, 170°W can be directly attributed to the retention of a bad observation. Combination of doubtful data into super-observations, and changes in the data selection sequence, have led to less substantial changes elsewhere. The positive impact of data validation procedures is undeniable and forms the cornerstone of the analysis project.

5. An analysis sequence: T_{150} for the 1986-87 El Niño

Figure 10 presents bimonthly analyses for the 150m depth-averaged temperature for the equatorial Pacific for the period 1986-87 which includes an El Niño event. The data are drawn from the TSDC files and analyzed on a 5° x 1° grid using Levitus' (1982) climatology as the first-guess field. TSDC have eliminated most of the poor data so the principal role of quality control here is to ensure the mapping is consistent with the imposed statistical structure. The event begins to take shape between the July/August and September/October 1986 bimonths. An unusual cold anomaly is located near 150°W on the equator. By the end of 1986 most of the central and eastern equatorial Pacific is anomalously warm. The cooler water just north of the NECC trough is probably an artifact of the excessive smoothing used in the climatology [Blomley et al. 1989]. The event matures by mid-1987 with eastern Pacific anomalies occasionally in excess of 4°C. The axis between the warm eastern Pacific and the cool western Pacific is coherent through most of 1987 being located near 165°W. Western Pacific cooling is of the same order as the eastern Pacific warming. By late 1987 cool water from the west can be seen penetrating toward the east along the equator signalling the decay of the event.

This example embodies the rationale for developing objective interpolation methods for the analysis of oceanographic data. Large data sets, spread unevenly in both space and time, can be efficiently mapped onto regular grids to facilitate better interpretation and utilisation of the data. There is however a clear need for a more refined treatment of statistical structure and for an improved first-guess field, either through the use of a ocean general circulation model (in a manner analogous to meteorology) or through stochastic methods (e.g. use of a persistence/regression model). These improvements are currently being examined.

6. References

Blomley, J.E, N.R. Smith, and G. Meyers, 1989. An oceanic subsurface thermal analysis scheme. Research Report, Bureau of Meteorology Research Centre .

Levitus, S., 1982. *Climatological Atlas of the World Ocean.* NOAA Prof. Paper 13, 173 pp., 17 microfiche, U.S. Govt. Printing Office, Washington, DC.

Lorenc, A., 1981. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701-721.

Meyers, G., J. Sprintall, H. Phillips and J. Peterson, 1989. Design of an ocean temperature observing network in the seas north of Australia. Part 1: Tropical Pacific ocean statistics. CSIRO Tech. Report.

White, W.B., G. Meyers, J.R. Donguy and S.E. Pazan, 1985. Short-term climatic variability in the thermal structure of the Pacific Ocean during 1979-82. *J. Phys. Oceanogr.*, **15**, 917-935.

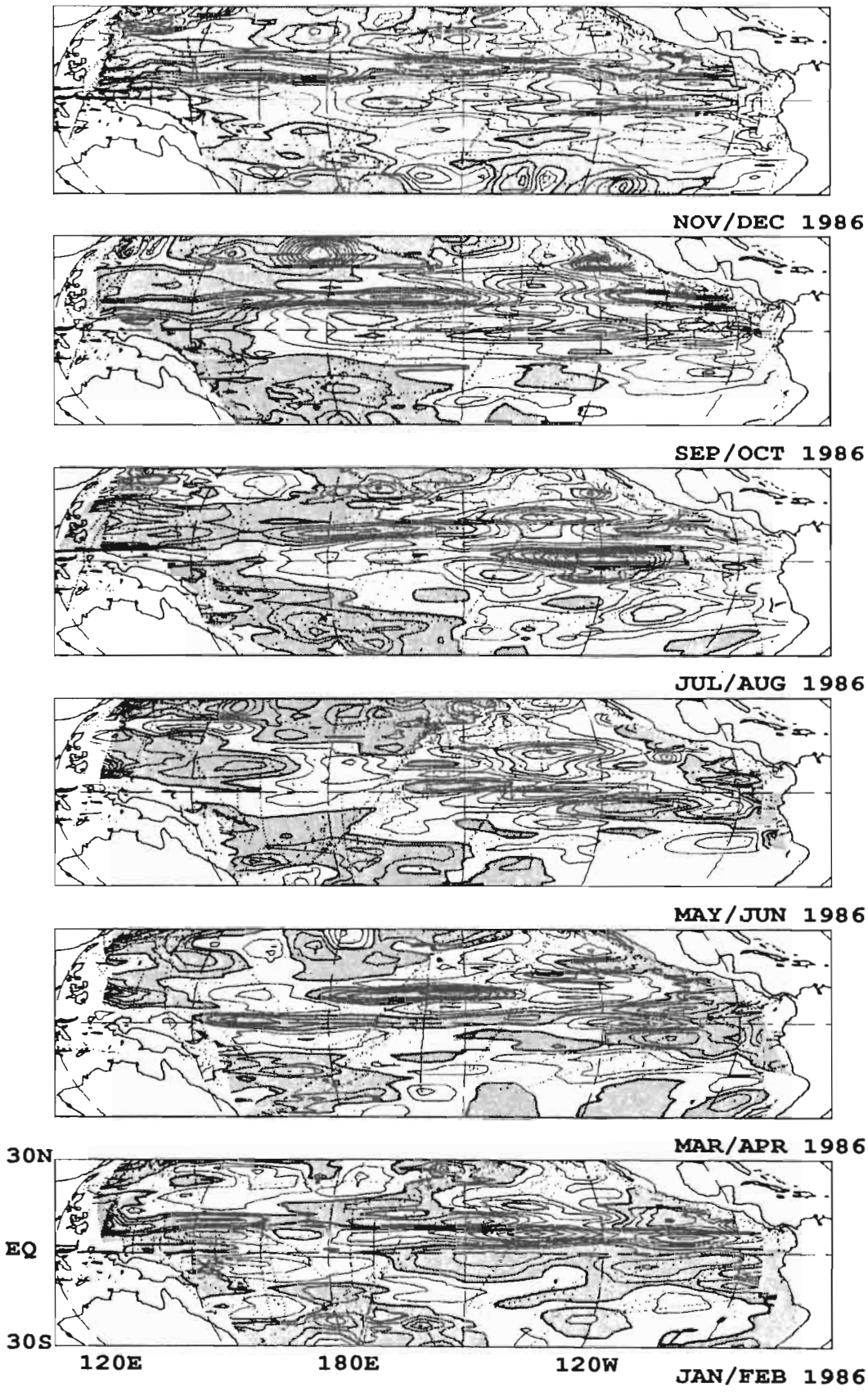


FIG. 10a. Bimonthly maps of the anomalous T_{130} field for 1986. Negative anomalies are shaded. The location dots mark the position of effective single and super-observations. The contour interval is 0.5°C with emphasised contours every 2°C.

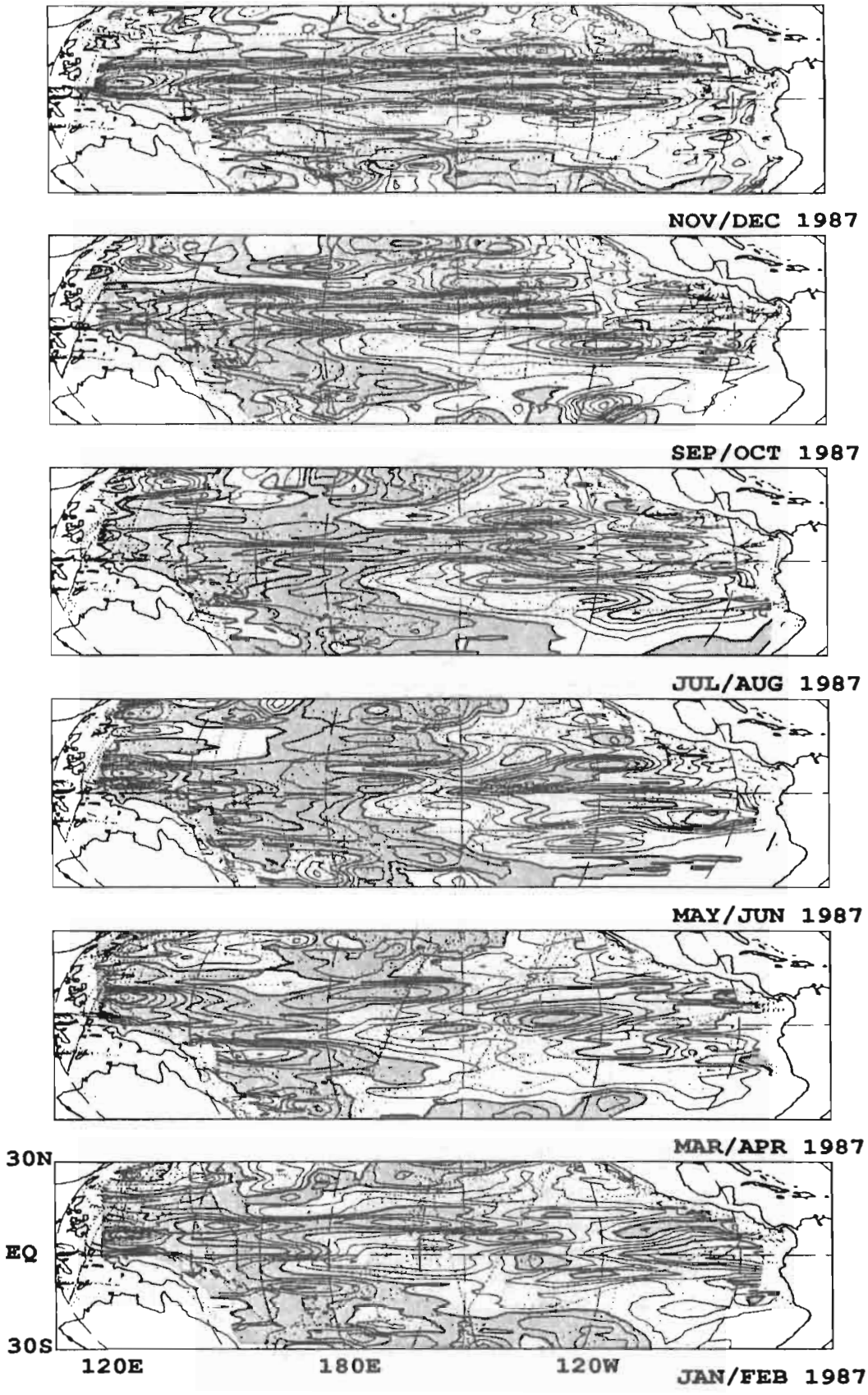


FIG. 10b. As Fig. 10a but for 1987.

WESTERN PACIFIC INTERNATIONAL MEETING AND WORKSHOP ON TOGA COARE

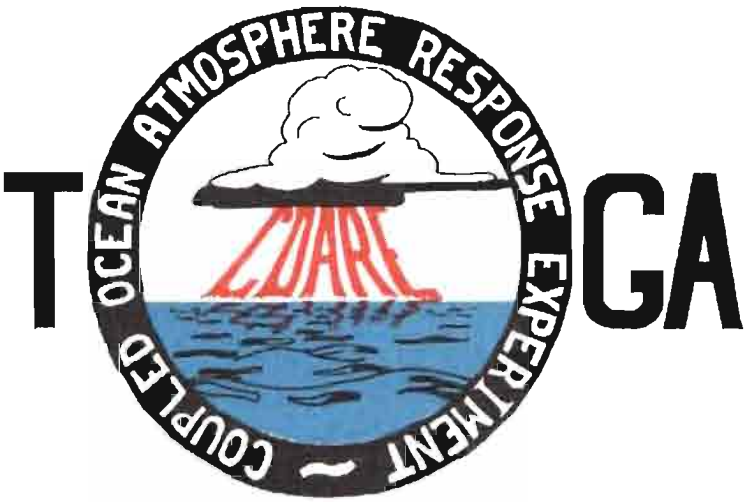
Nouméa, New Caledonia
May 24-30, 1989

PROCEEDINGS

edited by

Joël Picaut *
Roger Lukas **
Thierry Delcroix *

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



INSTITUT FRANÇAIS DE RECHERCHE SCIENTIFIQUE
POUR LE DÉVELOPPEMENT EN COOPÉRATION

ORSTOM

Centre de Nouméa

TABLE OF CONTENTS

MEETING COLLECTED PAPERS

WATER MASSES, SEA SURFACE TOPOGRAPHY, AND CIRCULATION

EL NINO/SOUTHERN OSCILLATION 1986-87

THEORETICAL AND MODELING STUDIES OF ENSO AND RELATED PROCESSES

MOMENTUM, HEAT, AND MOISTURE FLUXES BETWEEN ATMOSPHERE AND OCEAN

EMPIRICAL STUDIES OF ENSO AND SHORT-TERM CLIMATE VARIABILITY