Effect of subjective choices on the objective analysis of sea surface temperature data in the tropical Atlantic and Pacific oceans

Robert L. MOLINARI *, John F. FESTA

National Oceanic and Atmospheric Administration, Atlantic Oceanographic and Meteorological Laboratory, Physical Oceanography Division, 4301 Rickenbacker Causeway, Miami, Florida 33149, USA

Received 8 January 1999; revised 16 August 1999; accepted 20 September 1999

Abstract – Many subjective choices are required to perform an objective interpolation (OI) analysis of environmental variables. Herein, we consider the effects on the statistical analysis of sea surface temperature (SST) using (1) a structure function or covariance analysis, (2) different analytical expressions to represent the statistics of the raw data, and (3) different historical SST data sets. The historical data sets are the well-sampled Comprehensive Ocean–Atmospheric Data Set (COADS) and the poorly sampled historical expendable bathythermograph (XBT) data set. Results from these analyses are used to generate error maps for a poorly-sampled, two month XBT array and a proposed well-sampled profiling float array. For the relatively data-rich COADS analysis, decorrelation scales are the same using either the structure function or covariance analyses. Results differ for the data-poor XBT analysis. Representative decorrelation scales in the Pacific (Atlantic) are about 11–14 (6–10) degrees in the zonal direction and 4–7 (3–6) degrees in the meridional direction. As COADS SST data are less precise than XBT SST data, error and signal variances are greater for the former. The choice of analytical fit to the raw data (needed to generate error maps) has a dramatic effect on the resulting uncertainty fields. Gaussian fits, because of their parabolic shape near the origin, result in smaller errors than exponential fits for the same observing array. Finally, the proposed float array can achieve the accuracies needed to resolve satisfactory upper layer heat content changes over larger areas than the present XBT network. © 2000 Ifremer/CNRS/IRD/Éditions scientifiques et médicales Elsevier SAS

objective analysis / sea surface temperature / structure function / statistical analysis / objective interpolation

Résumé – Effet de choix subjectifs sur l'analyse objective des données de température superficielle dans les eaux tropicales des océans Altantique et Pacifique. Dans l'interpolation objective des variables environnementales, de nombreux choix sont subjectifs. Leurs effets sur l'analyse statistique de la température superficielle de l'eau (STT) sont comparés en utilisant : a) une fonction de structure ou analyse de covariance, b) des expressions analytiques des données brutes et c) des séries de données historiques. Celles-ci proviennent de séries complètes de données océan–atmosphère (COADS) et de séries fragmentaires de données de bathythermographes (XBT). Les résultats de ces analyses sont utilisés pour établir des cartes d'erreurs dans le cas d'un échantillonnage fragmentaire par réseau de bathythermographes sur une durée de deux mois et dans le cas du réseau de flotteurs proposé pour l'enregistrement des profils. Pour les séries COADS relativement riches en données, les échelles de décorrélation sont les mêmes dans les analyses par fonction de structure ou convariance. Les résultats diffèrent pour les séries XBT fragmentaires ; les échelles de décorrélation dans le Pacifique sont environ 11–14° dans la direction zonale et 4–7° dans la direction méridienne ; dans l'Atlantique, elles sont respectivement de 6–10° et 3–6°. Les données COADS étant moins précises que les données XBT, les variances d'erreur et de signal sont

* Correspondence and reprints: molinari@aoml.noaa.gov

PII: S0399-1784(00)00108-0/FLA

plus grandes pour les premières. Le choix de l'ajustement analytique pour les données brutes (nécessaire pour établir les cartes d'erreur) a un effet important sur les champs d'incertitude qui en résultent. Les ajustements gaussiens, par leur forme parabolique au voisinage de l'origine, donnent des erreurs inférieures à celles des ajustements exponentiels pour un même réseau d'observation. Pour résoudre la variabilité des échanges de chaleur dans la couche superficielle, le réseau de flotteurs proposé donne une précision satisfaisante sur des zones plus étendues que l'actuel réseau de bathythermographes.

analyse objective / température superficielle / fonction de structure / analyse statistique / interpolation objective

1. INTRODUCTION

Upper ocean temperature (UOT) data are an important component of many oceanographic studies (e.g. the World Ocean Circulation Experiment (WOCE), WMO [18], the Tropical Ocean Global Atmosphere (TOGA) experiment, NRC [12]). Objectives for these data include initializing climate forecast models and estimating upper layer heat content changes. The majority of the data collected for these programs have been obtained from a volunteer observing ship (VOS) network. The VOS network is comprised of merchant ships that typically have fixed routes on which ship's crew members deploy expendable bathythermographs (XBTs).

Objective interpolation (OI) schemes are commonly used to map onto a fixed grid a non-uniform distribution of temperature data from the VOS network and to design 'optimal' sampling networks for the collection of these data. For both the mapping and design functions, an accurate representation of the statistics (i.e. noise and signal variances and decorrelation scales) of the variable to be mapped is an essential requirement of OI schemes. In mapping, these statistics are required to define the coefficients for an optimal linear predictor such that the mean square error between the unknown true value and the estimated value is minimized. Once the statistics are determined, gridded values can be predicted and a measure of the reliability, in the form of an error map, can also be calculated. In design studies, both the variability statistics (i.e. extracted from existing data) used in mapping and various sampling strategies (i.e. the locations of proposed observations) are needed. Once the statistics are determined, the actual values of the data are not required to calculate the uncertainty estimates. Error fields are generated for a variety of sampling patterns. An 'optimal' array is determined subjectively based on operational costs and accuracies desired.

There are several important subjective choices required in OI for both mapping and design. The effect of these choices on the OI results is frequently amplified in oceanographic relative to meteorological applications because of the paucity of the historical data available to generate the statistics needed to map the data. For example, an initial choice in OI is the form of the statistical analysis used to represent the characteristics of the field to be mapped. Historically, the analysis of oceanographic data was typically based on a correlation (covariance) analysis to estimate scales and variances from historical data (e.g. White [15], Meyers et al. [11], and Sprintall and Meyers [14]). Recently, structure functions have been used to represent the statistics of the variability, e.g., Festa and Molinari [5] and Hansen and Herman [7]). When adequate data are available to obtain reliable statistics, an analysis using either a structure or correlation function will give the same estimates of variability (Cressie [2], Gandin [6], and Herzfeld [8]). However, there are cases (e.g., Hansen and Herman [7]) where it is difficult if not impossible to estimate a covariance function and the statistical representation can only be given in terms of a structure function (e.g. a linear structure function).

In the application of OI to oceanographic problems, analytical expressions are fitted to the raw autocovariance or structure function representations of the historical data to provide the quantification of the properties of the variability needed for mapping. The fits are required as the raw data representations are frequently noisy because of sparse data, and look-up tables of these data may not possess the properties needed to satisfy theoretical OI requirements (i.e. the fit must be positive definite to satisfy OI theory, Isaaks and Skivastav [9]). Considerable subjectivity is involved in the selection of a suitable analytical function (e.g. White et al. [16] and Meyers et al. [11] employed Gaussian functions to represent spatial statistics, while using similar data sets, Festa and Molinari [5] and White [15] used exponential functions). Using a Gaussian function instead of an exponential function can result in the reporting of significantly lower error estimates because of the shape of the Gaussian function at short lags (Isaaks and Skivastav [9]).

Finally, different oceanographic data sets are obtained using instruments with different accuracies and precisions. For example, SST from VOS use injection and bucket instruments. However, the depth of the ship's intake varies considerably, causing sampling of different levels of the near surface. Bucket temperatures are sensitive to the construction of the instrument. Thus, SST from VOS are generally less precise than those from, for example, XBTs (i.e. a single observing method).

Herein, we consider the effect on OI results of using (1) a structure function or covariance analysis, (2) different historical data sets, and (3) different analytical expressions to represent the statistics of the raw data. We begin with a brief description of the analytical method, followed by a discussion of the sources of the historical data. Examples are then presented, using actual and proposed observational locations, to illustrate the effect of different analysis choices on the characteristics of the resulting error maps. We conclude with a summary and a discussion of the analysis of these data sets and their implications for design studies.

2. METHOD OF ANALYSIS

Gandin [6] initially developed the statistics of the variable (Z) to be mapped in terms of the structure function S_{ij} (i.e. the 'known' statistics) defined as

$$S_{ij} = S(x_i, x_j) = \langle [Z(x_i) - Z(x_j)]^2 \rangle \quad (1)$$

the mean-square difference between the values of Z at locations x_i and x_j. The structure function, or semivariogram (1/2 S_{ij}), is determined from previous observations of the variable considered. The value S_{ij} can be represented by $S(h)$ where h is the lag, a measure of the distance between variable pairs. There is a relationship between the structure function $S(h)$ and the autocorrelation function $C(h)$, when both are well defined (see Gandin [6]), such that

$$C(h) = 0.5^*[S(\infty) - S(h)] \quad (2)$$

Perhaps the most important step in either a structure function or covariance analysis is the selection of an analytical expression to approximate the statistics of the raw data. In structure function analysis, this expression is known as the semivariogram. When semivariograms are only dependent on the magnitude of the lag, the statistics are said to be isotropic. Two commonly used isotropic forms are the exponential and Gaussian structure functions written as

$$\text{Exponential:} \quad S(h) = C_0 + C_1{}^*(1 - \exp(-h/C_2)), \quad (3a)$$

$$\text{Gaussian:} \quad S(h) = C_0 + C_1{}^*(1 - \exp(-h/C_2)^2). \quad (3b)$$

The value of the semivariogram at lag zero, C_0, is referred to as the nugget, which is a measure of both the instrument and subgrid geophysical noise. In these functional forms, the asymptotic value of the semivariogram, $C_0 + C_1$, is known as the sill and represents the total variance (noise and signal) of the field. The value C_2 is a measure of the range over which correlation exists and is equivalent to the e-folding scale of the covariance function. The practical range, i.e. the distance at which the correlation function effectively has reached zero, is given as 3^*C_2 for the exponential and $\sqrt{3}^*C_2$ for the Gaussian function. The exponential model shows a linear behavior near the origin, while the Gaussian model shows a parabolic behavior.

When the variability of the semivariograms for two-dimensional spatial variables (i.e. the sill and range) depends on both magnitude and direction of the lag, the statistics are said to be anisotropic. Anisotropies typically result from the underlying physical processes evolving differently in space. If an anisotropic functional form (one with the same component sills but different ranges) can be reduced to isotropic form by a simple linear transformation of coordinates, the anisotropy is referred to as geometric; otherwise, it is known as zonal.

Figure 1 shows examples of east–west and north–south isotropic and anisotropic semivariogram functions. The semivariograms are the same in each direction (i.e. they have the same sill and range) in the isotropic example. In the geometric anisotropic example, the sill, i.e. total variance, is the same in each direction and the range in the north–south direction is half of that in the east–west direction. A simple linear transformation can reduce the semivariograms to isotropic functions. In the zonal anisotropic example, the variance in the north–south direction is 50 % higher than in the east–west direction. It is not possible to reduce these variograms to isotropic functions.

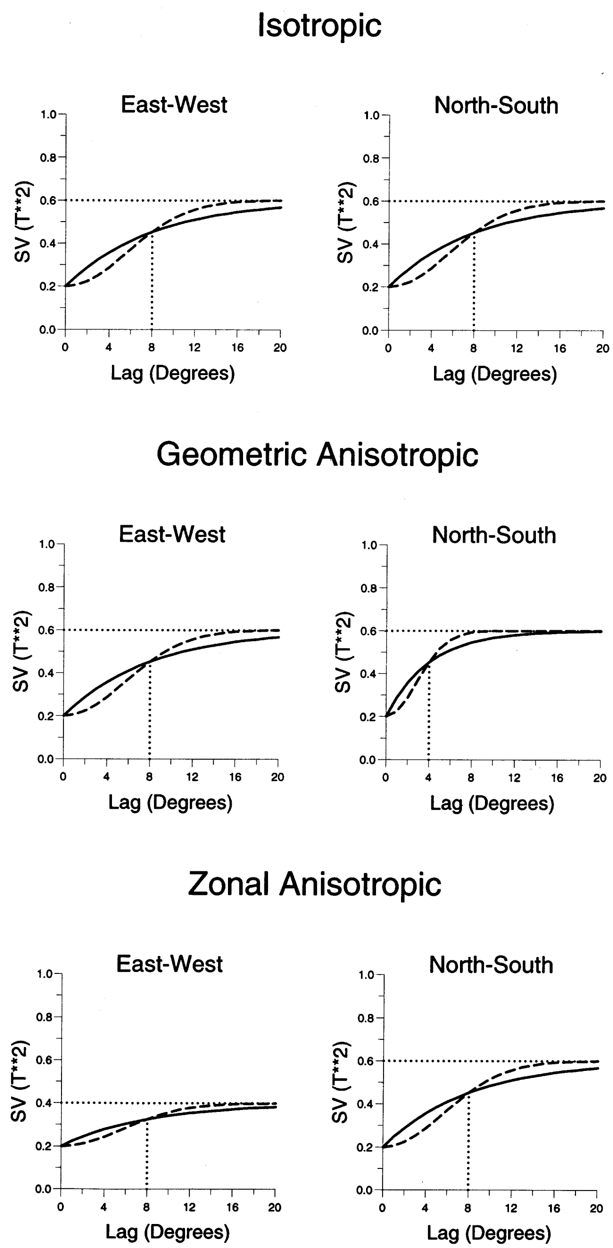

Figure 1. Model structure functions using exponential (solid lines) and Gaussian (dashed lines) functions for the east–west (left panels) and north–south (right panels) directions. Isotropic (upper panels), geometric anisotropic (middle panels), and zonal anisotropic (lower panels) analytical forms are shown. Sills and ranges of the fits are given by dotted lines.

3. DATA SOURCES AND ANALYSIS

The highly sampled COADS SST data and the poorly sampled XBT data collected in the tropical Atlantic and Pacific Oceans are used in this study. It is recognized that XBT data are not ideal for performing SST analysis because of limited spatial coverage. However, they are adequate for the purposes of this study. The COADS SST data are comprised of measurements collected and reported, along with surface meteorological observations, by merchant ships. We use the monthly summary data on a 2 degree latitude by 2 degree longitude grid for the time period of January 1950 through December 1989 (see Woodruff et al. [17] for details of the data and binning procedures).

The summary SST data contains both mean and median values. We have chosen to present the 40-year (1950–1989) median values analysis, since the median is less likely to be affected by questionable data values. Analysis has also been performed using mean values and over shorter decadal time periods (i.e. 1950–1959, ..., 1980–1989). There is no significant difference in the results. Results are also similar between the 40-year and decadal time periods; in the Pacific Ocean this occurs because of a uniform distribution of El Niño and La Niña events during the decadal time periods.

The tropical Atlantic and Pacific Oceans between 30° S and 30° N, the former between 70° W and 15° E and the latter between 150° E and 90° W, are considered in this study. For the COADS data, during the period of January 1950 through December 1989, long-term monthly means are calculated at each 2° by 2° grid node where there are at least five years of data. Monthly anomaly fields are obtained by subtracting the long-term monthly mean climatology from the yearly values at each grid node.

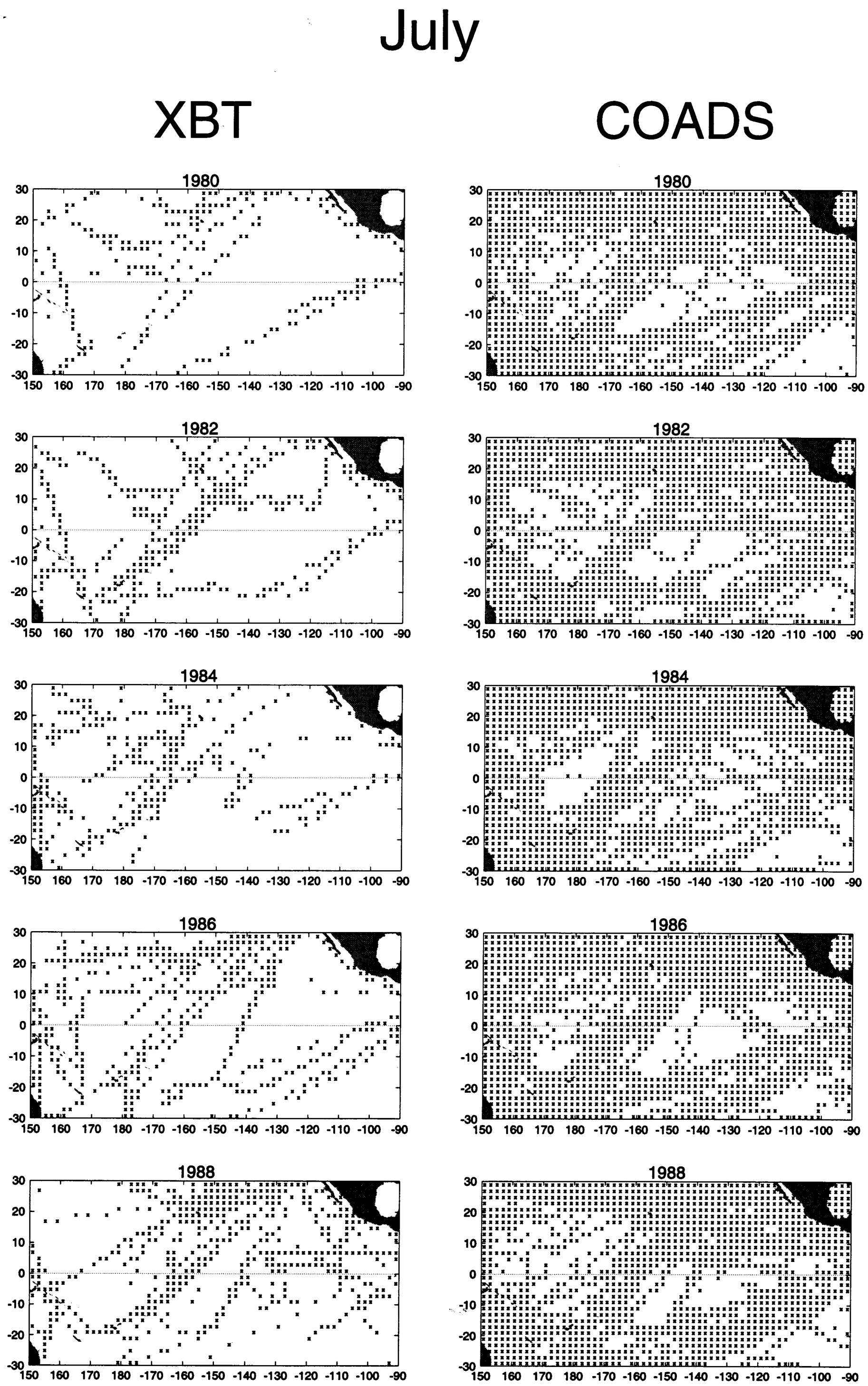

Figure 2. July XBT (left panels) and COADS (right panels) data distributions in the tropical Pacific for even-numbered years during the period 1980–1988.

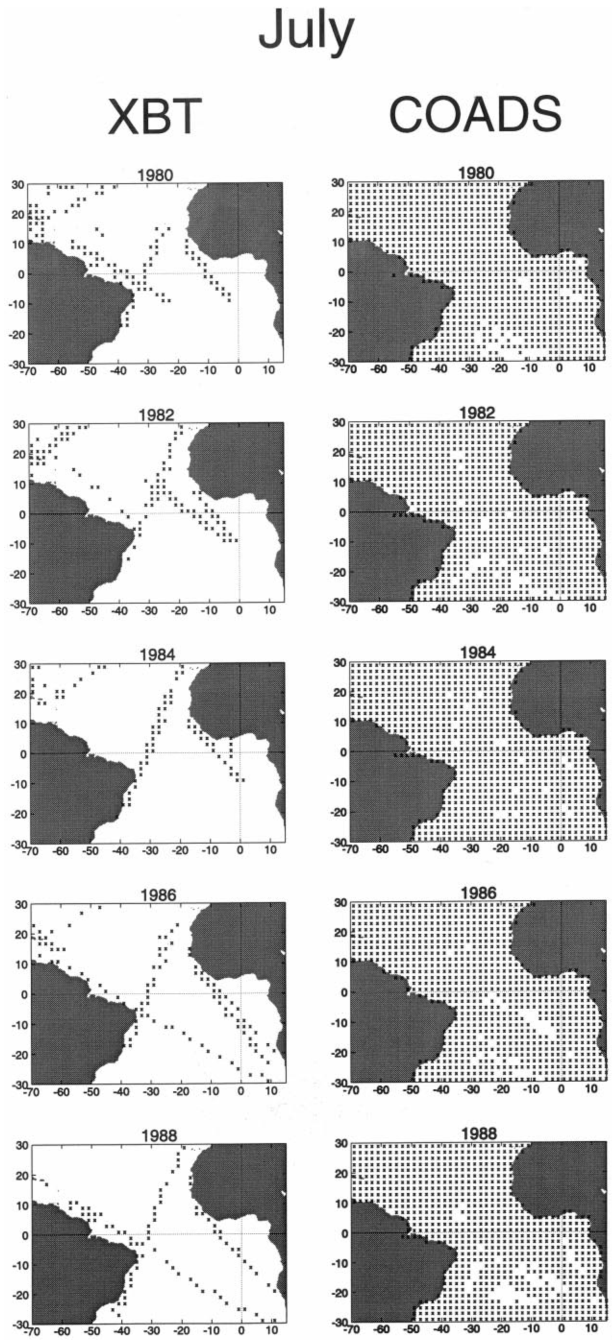

Figure 3. Same as *figure 2*, except for the tropical Atlantic.

The historical XBT data have been compiled for the Atlantic and Pacific Oceans within the same tropical regions. In the Atlantic, the historical XBT data have been quality controlled for the period of June 1966 through December 1991 using procedures similar to those given in Daneshzadeh et al. [3] and in the Pacific for the period of January 1979 through December 1993, Donoso et al. [4]. The XBT data are averaged by month and year onto the COADS 2° by 2° data grid. Long-term means and anomalies are calculated using the same approach as applied to the COADS data. Examples of the XBT and COADS data distribution on this grid for the month of July are given in *figure 2* for the Pacific and *figure 3* for the Atlantic. The dramatic disparity between XBT and COADS data availability, particularly in the later years, is evident in the tropical Atlantic.

Mean squared difference values (i.e., structure functions or semivariograms) in the north–south (meridional) and east–west (zonal) direction were computed for each 2 degree lag bin for each month and each year with data. Monthly composites were formed by taking the average value at any bin over the entire time period of record. Seasonal (December–February, March–May, June–August, and September–November) and annual composites were calculated in a similar manner. There is little seasonal variability in the COADS semivariograms in either tropical ocean (not shown). More variability, especially in the Atlantic, is evident in the XBT semivariograms at all lags, which is the result of the lack of sufficient data during any seasonal time period. Thus, further discussion is limited to the annual composite semivariograms.

Analytical functional forms are fitted to the binned values using the following approach. First, the noise variance (nugget) is calculated in each direction by fitting only the first few bins through a simple linear extrapolation to zero lag, Alaka and Elvander [1]. An average nugget (C_0) is then computed from these values. The signal variance and decorrelation scales are then determined by a fitting routine using the previously estimated nugget. To arrive at 'best-fit parameters', the signal variances (C_1) are estimated from a review of the raw structure function and e-folding scales (C_2) are then calculated using these values in the fitting scheme. This approach satisfies the geostatisticians' warning against blind automatic fitting of parameters to the raw semivariograms (see Journel and Huijbregts [10], for example).

Two realizations of data distribution are used. First, XBT data collected and transmitted in real-time during January and February 1998 by NOAA's VOS program are assumed representative of a typical two-month period. Additional XBT data are generally available from non-NOAA sources but frequently not in real-time. Second, an Array for Real-time Geostrophic Oceanography (ARGO) has been proposed, Smith [13]. ARGO includes a global deployment of Profiling ALACE (PALACE) floats. These floats drift at a preselected depth (generally between 1 000 m and 1 500 m) for some 10 to 15 days. They return to the surface collecting profiles of temperature (and salinity, if equipped with a conductivity sensor). At the surface they transmit their position and profile data. Initial plans call for maintaining a global deployment of about 3 000 floats. Such an array will provide about 300-km resolution of upper layer properties. A random distribution of floats with this average spacing is used to show the effect of various OI choices on ARGO error fields.

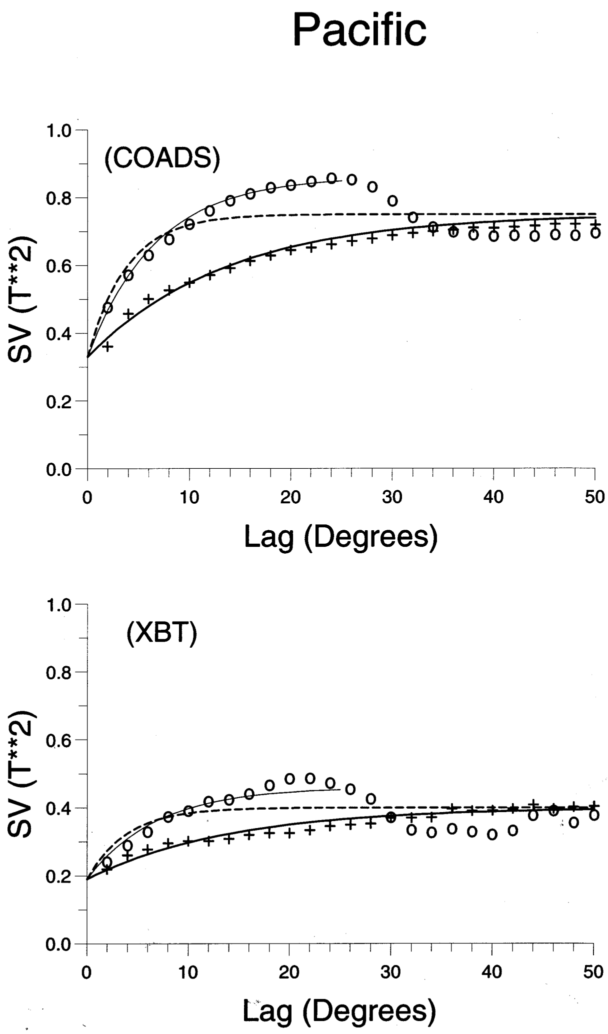

Figure 4. Annual composite SST structure functions for the COADS (upper panel) and XBT data (lower panel) generated from raw anomaly data in the tropical Pacific. The heavy solid and dashed lines represent geometric anisotropic analytical fits to the east–west (zonal) and north–south (meridional) direction, respectively, computed using 25 lags (50°). The light solid line represents a zonal anisotropic fit in the north–south direction using 13 lags (26°).

4. RESULTS

4.1. Estimates of noise and signal variances and decorrelation scales

Raw structure function values estimated from the Pacific COADS and XBT SST data sets are shown in *figure 4* and from the Atlantic data in *figure 5*. For illustrative purposes, the raw values in the zonal (+) and meridional (O) directions are plotted to 50 degrees (25 lags). However, the noise in the Atlantic XBT structure function at longer lags (*figure 5*) argues for limiting analysis of these data to separations less than about 26 degrees (13 lags).

The choice of data set has an effect on the resulting noise and signal variances. Variances associated with the COADS data are larger than the XBT variances in both basins (*figures 4*, *5*). SST observations from XBT data are expected to be more precise than those from VOS because of the many different measurement techniques (e.g., bucket, intake, etc.) and sampling depths (dependent on intake depth of the VOS) used on the merchant ships. This lower precision contributes to the higher variances estimated from the COADS data and would contribute to larger uncertainties when mapping SST data from VOS.

The Pacific meridional structure functions both show the presence of what geostatisticians call a 'hole' (i.e., a decrease rather than increase with lag in the values of the variogram, *figure 4*). Geophysically, the hole structure, if real, is representative of solutions to second order autoregressive processes (i.e., the processes include periodic components, White [15], for example). Anomaly plots constructed for the basin, particularly during ENSO events, suggest that the

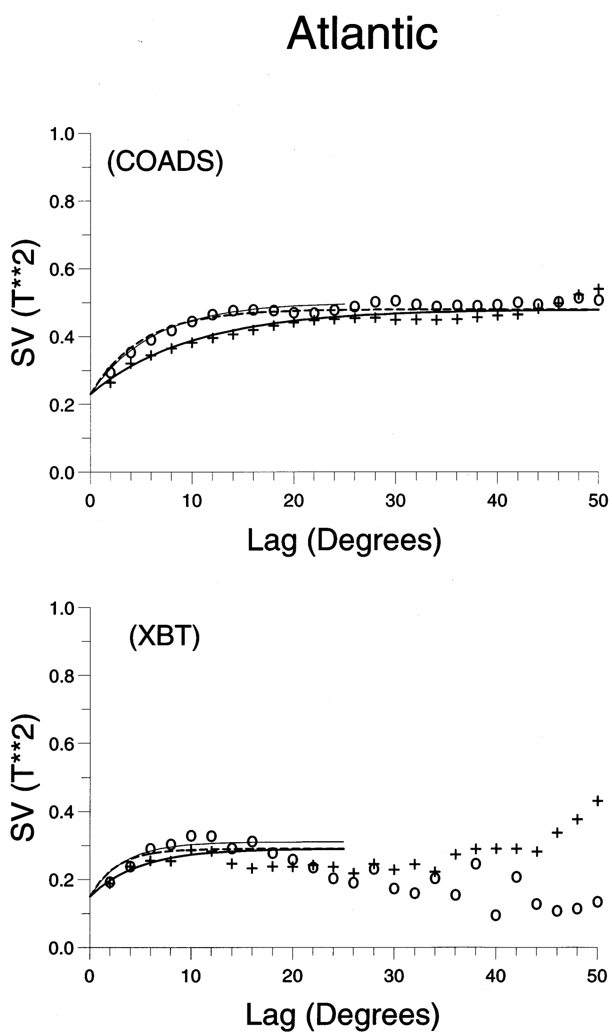

Figure 5. Same as *figure 4*, except for the Atlantic and using only 13 lags (26°) for the XBT data.

structure is real. The ENSO anomaly pattern is characterized by an anomaly pattern of one sign extending some 15–20 degrees on either side of the equator bounded by anomalies of another sign. This anomaly structure causes higher variances at shorter lags (order 15 degrees) than longer lags (greater than 20 degrees).

The hole structure complicates the fitting of analytical forms to the raw variograms as the characteristics of the fit, in particular, the signal variances, become strongly dependent on the number of lags used in this process. A common approach to fit these types of raw structure functions is to assume geometric anisotropy, as at the longest lags the sills are approximately equal. The Atlantic XBT data are noisy at the longest lags as noted above, so this approach is not valid for these structure functions. A representative sill is selected from the raw structure functions and with the estimated nugget, decorrelation scales are estimated from a fitting routine. The COADS and XBT raw structure functions are best represented by exponential fits (*figures 4, 5*). The estimated properties of these fits are given in *table I*.

In the Pacific, the decorrelation scales from both the COADS and XBT data are very similar. White [15] used correlation functions to estimate decorrelation scales from the same XBT data set but with a somewhat different preconditioning applied to the raw data using different averaging intervals. However, his Pacific scales are very similar to those estimated using structure functions. In addition, the variances estimated from the XBT data indicate a signal to noise ratio of approximately one, again similar to White's findings. In the Atlantic, the XBT and COADS scales are different by a factor of two, with White's [15] estimates falling between the structure function estimates. A portion of these differences can be explained by the use of fewer lags with the XBT data because of the noise at the longer lags (*figure 5*).

When fitting over longer lags (i.e., over the 'hole effect' in the Pacific), in order to maintain isotropic behavior through a geometric transformation, the meridional variances are underestimated at the shorter lags (*figures 4* and *5*). Alternatively, the structure functions could be fitted using fewer lags, thereby representing the data more faithfully. Using the latter approach and the Pacific XBT data, meridional signal variances for an exponential fit range from $(.32\ °C)^2$ to $(.22\ °C)^2$ and zonal variances from $(.14\ °C)^2$ to $(.21\ °C)^2$ when lags range between 10 and 20, respectively. Similarly, in the Atlantic, the differences in variances for a similar fit and lag choices are $(.07\ °C)^2$ in the meridional direction and negligible in the zonal direction. The effect of different variances on error maps will be demonstrated in the next section.

Table I. Properties (C_0, C_1, and C_2) of analytical fits to the raw structure functions shown in *figures 4* and *5* using exponential forms (equation 3a) and assuming geometric anisotropy. C_0 is the nugget, C_1 the sill (or signal variance) and C_2 the range (or decorrelation scale). N is the number of 2 degree lags used in the fitting routine (i.e. 20 lags = 40 degrees). Values for the zonal anisotropic structure function are also presented. The White data set decorrelation scales (C_2) are from [15]. The units for C_0 are $(°C)^2$, C_1 $(°C)^2$, and C_2 (degrees). GA and ZA correspond to the geometric anisotropic and zonal anisotropic values, respectively.

Data Set	N	C_0	C_1 (Zonal)	C_1 (Meridional)	C_2 (Zonal)	C_2 (Meridional)
Pacific:						
COADS (GA)	20	0.33	0.42	0.42	13.7	3.8
XBT (GA)	20	0.19	0.21	0.21	13.9	3.9
COADS (ZA)	13	0.33	0.42	0.52	13.4	6.6
XBT (ZA)	13	0.19	0.21	0.27	14.1	6.4
White					11.2	5.6
Atlantic:						
COADS (GA)	20	0.23	0.25	0.25	10.0	5.0
XBT (GA)	13	0.15	0.14	0.14	5.5	2.6
COADS (ZA)	13	0.23	0.25	0.27	9.8	6.2
XBT (ZA)	13	0.15	0.14	0.16	5.5	3.5
White					7.0	3.0

4.2. Error analysis

The effect of fitting choices on uncertainty estimates is demonstrated by constructing error maps from XBT data collected during January and February 1998 (a two-month interval is necessary because of data sparsity during a one-month interval) and a representative ARGO-PALACE float array. The ARGO array demonstrates the effect of the choices on a well-sampled region and the XBT array, on a poorly-sampled area. If the WOCE objective of providing constraints on surface heat flux estimates is chosen as a rationale for collection of the UOT data, desired accuracies in the resulting fields can be derived. To obtain a 15 W/m^2 accuracy in heat content changes over a two-month interval and 40-m-thick layer requires temperature uncertainties less than 0.5 °C. Thus, on the error maps generated, uncertainties less than 0.5 °C are shaded.

Using the representative variances and scales derived from the XBT data set (*table I*) and a geometric anisotropic exponential fit, it is seen that in the Pacific the ARGO data distribution can achieve accuracies of less than 0.55 °C (17 W/m^2) over the entire basin, with many regions attaining the desired 0.5 °C (15 W/m^2) accuracy (*figure 6*). Using the same fit, the sparse XBT data distribution includes large areas where interpolation is not even possible. Desired accuracies are only achieved in regions with dense sampling along lines and/or contiguous transects (*figure 6*).

White [15] estimated average zero-crossing decorrelation scales for 10° bands in both basins. Averages for the band 30° S to 30° N are given in *table I*. However, because of the large variability in scales in a particular basin, White [15] goes on to suggest that minimum scales (5.0 in the zonal direction and 2.5 in the meridional direction) resulting from his analysis of both surface and subsurface data should be employed in OI mapping exercises. The shorter scales degrade the accuracies for both the ARGO and XBT data distributions (*figure 6*). Because of the shape of the Gaussian fit near the origin (*figure 1*), the 0.5 °C criteria is realized throughout the Pacific using the ARGO data distribution.

As indicated previously, because of the shape of the meridional structure function, the level of the sill is strongly dependent on the number of lags used to fit the raw Pacific data (*figure 4*). As a worst case, a nugget of (0.38 °C)^2 and a signal variance of (0.52 °C)^2 were obtained from fitting the Pacific COADS meridional structure functions with an exponential fit. Error maps using these estimates and decorrelation scales from *table I* show typical uncertainties of about (0.75 °C)^2, 22 W/m^2, for the ARGO data throughout the basin and the same for the XBT data along the tracklines (*figure 7*).

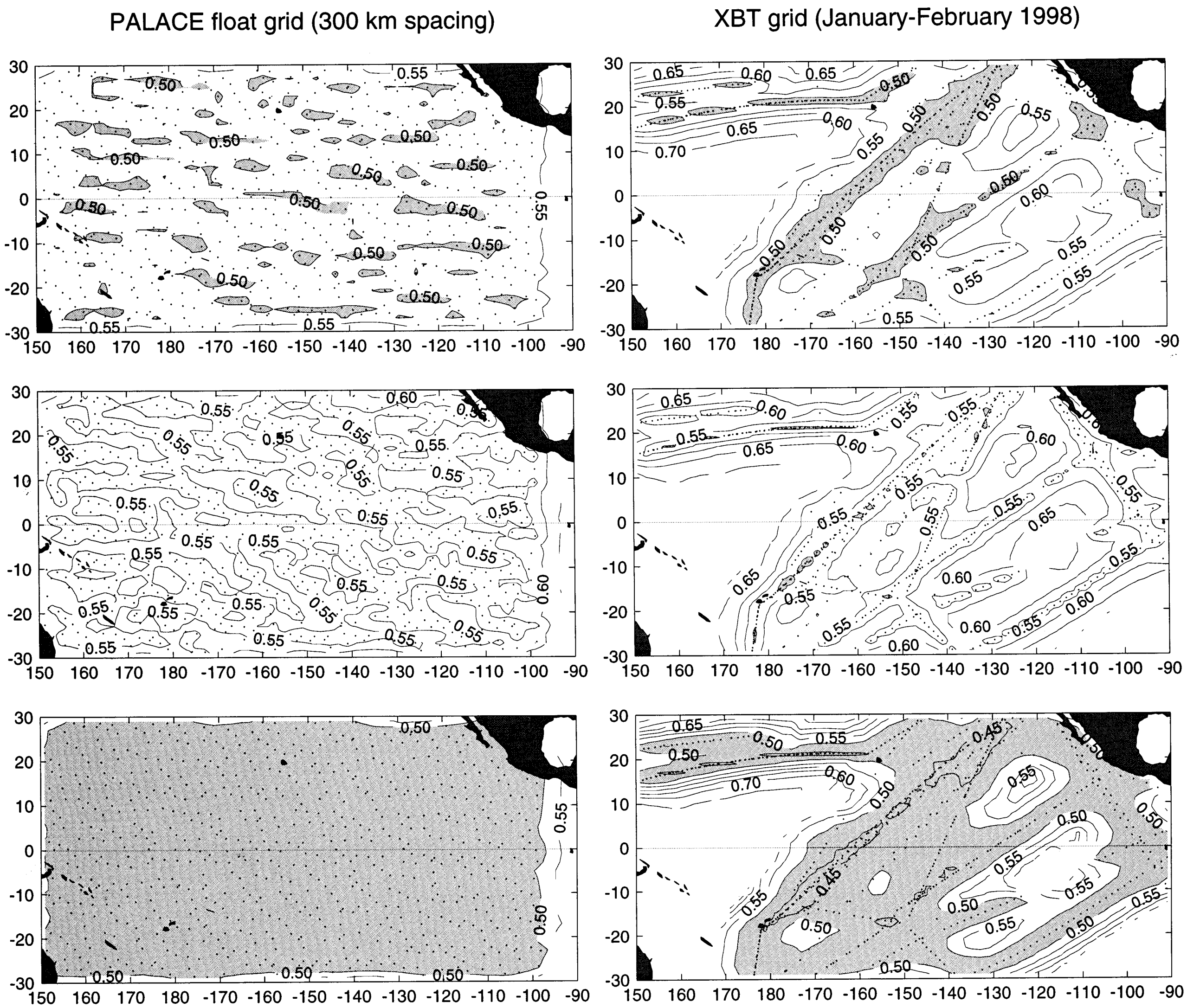

Figure 6. Pacific error fields (°C) for three cases: Upper panels: A 300-km spacing ARGO PALACE float grid (left panel) and the January–February 1998 real-time XBT grid (right panel). Geometric anisotropic exponential analytical fits to the raw data (fit parameters are given in *table I*) were used to generate the error fields. Middle panels: same as upper panels, except for the use of White's [15] minimal decorrelation scales (zonal, 5°, meridional, 2.5°). Lower panels: same as upper panel except for the use of a Gaussian anisotropic analytical expression in the fitting of the raw structure functions.

The nuggets and signal variances as well as the decorrelation scales estimated from the Atlantic XBT data are less than those estimated from the Pacific data (*table I*). Error maps have been generated using the representative Atlantic values. The desired accuracy of 0.5 °C is achieved throughout the tropical Atlantic if ARGO spatial resolution is available and within the area bounded by the tracklines if XBT data are used (*figure 8*).

5. SUMMARY AND CONCLUSIONS

The noise and signal variances estimated from the COADS data are greater than those generated from the XBT data (order 0.2 °C, 6 W/m^2, in the Pacific, *table I*). Thus, care must be taken when combining VOS and XBT data in a mapping exercise. The higher COADS variances should be used to obtain more realistic error fields if both data are used.

The results of the analysis of XBT data by White [15] and that presented above are similar in the tropical Pacific, both giving signal-to-noise ratios of about 1 and representative decorrelation scales of 11–14 degrees in the zonal direction and 4–7 degrees in the meridional direction (*table I*). Thus, when adequate data are available either a structure or correlation function representation of the raw data is appropriate.

At the sea surface the XBT data in the Atlantic give decorrelation scales that are 50 % less than those estimated from the COADS data and given by White [15]. The noise in the XBT structure functions at longer lags (*figure 5*) precludes using these data in the analytical fitting operation, contributing to the shorter scales. The White [16] and COADS results suggest that for the Atlantic, representative decorrelation scales in the zonal direction are of the order 7–10 degrees and in the meridional direction 3–6 degrees (*table I*).

White [15] develops reasonable decorrelation scales in the Atlantic using approximately the same XBT data set as here but using different preconditioning of the data (he uses 5 degrees of longitude by 2.5 degrees of latitude bins, which limits the minimal scales he can resolve). We conclude that preconditioning is an important step in the OI procedure. The resulting limits in resolution caused by using larger averaging bins must be considered in the mapping and design exercises that use the results.

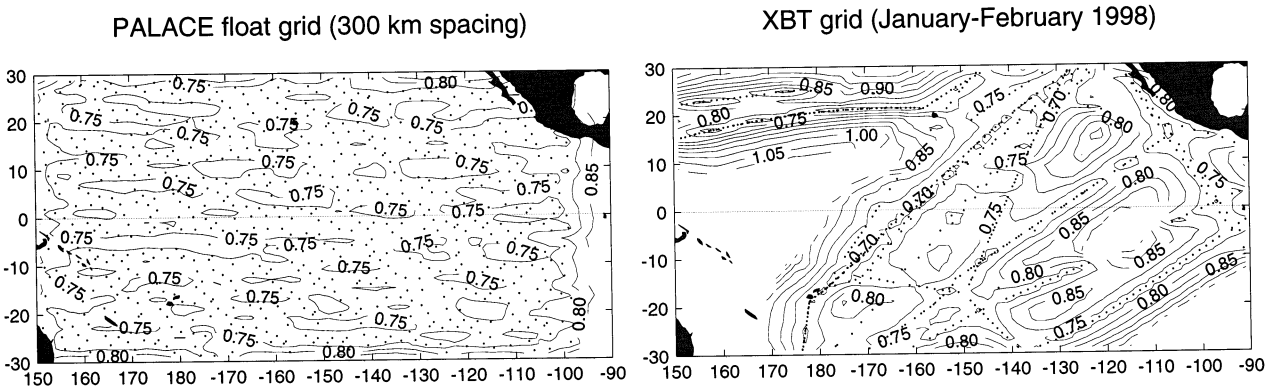

Figure 7. Same as *figure 6*, upper panel, except for the use of a larger nugget and signal variance (see text).

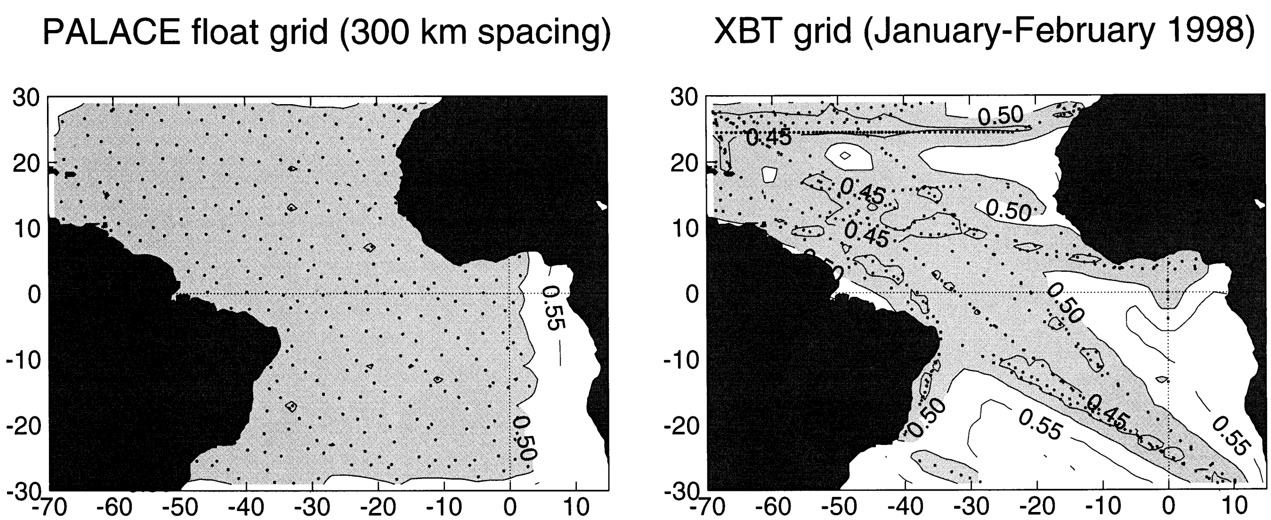

Figure 8. Same as *figure 6*, upper panel, except for the tropical Atlantic.

For the same data set and form of isotropy, there is little difference in the scales resulting from either an exponential or Gaussian fit (not shown). The latter fit does provide smaller error estimates because of the shape of the function at the shortest lags (*figure 1*). The raw XBT data are better represented by the exponential function (*figures 4* and *5*), although others use the Gaussian function for the same data [11].

The use of minimal scales as proposed by White [15] would result in similar error fields to those obtained from using the longer scales in *table I* (*figure 6*). As stated previously, it is recognized that XBT data are not the ideal observation to generate global SST distributions. However, White [15] estimates that spatial scales vary little at least to 200 m. Thus, using the XBT distributions to map near surface temperature distributions with a similar accuracy constraint would require considerably greater coverage (and expense) than is available with the present network. For instance, using the scale analysis frequently quoted in OI studies that two to three samples are required per decorrelation scale, using the minimum scales of 5° by 2.5° rather than the 13.0° by 5.0° scales in the tropical Pacific would, at a minimum, more than double the number of probes required.

REFERENCES

[1] Alaka M.A., Elvander R.C., Optimum interpolation from observations of mixed quality, Mon. Wea. Rev. 100 (1972) 612–624.

[2] Cressie N., The origins of kriging, Math. Geol. 22 (1990) 239–252.

[3] Daneshzadeh Y.-H., Festa J.F., Minton S.M., Procedures used at NOAA-AOML to quality control real-time XBT data collected in the Atlantic Ocean, NOAA Tech. Memo., 1994, ERL AOML-78, 44 p.

[4] Donoso M.C., Harris J.E., Enfield D.B., Upper ocean thermal structure of the eastern tropical Pacific, NOAA Tech. Rep., 1994, ERL 450-AOML 36, 221 p.

[5] Festa J.F., Molinari R.L., An evaluation of the WOCE volunteer observing ship XBT network in the Atlantic Ocean, J. Atmos. Oceanic Tech. 9 (1992) 305–317.

[6] Gandin L.S., Objective Analysis of Meteorological Fields, Gidrometeorizdat, 1963, 266 p. (reprinted in 1965 by the Israel Program for Scientific Translations, Jerusalem, 238 p.)

[7] Hansen D.V., Herman A., Evolution of isotherm depth anomalies in the eastern tropical Pacific during the El Niño event of 1982–1983, J. Geophys. Res. 94 (1989) 14 461–14 473.

[8] Herzfeld U.C., Inverse theory in the earth sciences: An introductory overview with emphasis on Gandin's method of optimum interpolation, Math. Geol. 28 (1996) 137–160.

[9] Isaaks, E.H., Skivastav R.M., Applied Geostatistics, Oxford Univ. Press, 1989, 600 p.

[10] Journel A.G., Huijbregts C.J., Mining Geostatistics, Academic Press, New York, 1978, 600 p.

[11] Meyers G., Phillips H., Smith N., Sprintall J., Space and time scales for optimal interpolation of temperature: Tropical Pacific Ocean, Prog. Oceanogr. 28 (1991) 198–218.

[12] NRC, Ocean–Atmosphere observations supporting short-term climate predictions: panel on near-term development of operational ocean observations, National Academy Press, 1994, 31 p.

[13] Smith N., Prospectus for ARGO: A global profiling float array, Exchanges (newsletter of the Climate Variability and Predictability Programme, CLIVAR) 3 (1998) 15–19.

[14] Sprintall J., Meyers G., An optimal XBT sampling network for the eastern Pacific Ocean, J. Geophys. Res. 96 (1991) 10 539–10 552.

[15] White W.B., Design of a global observing system for gyre-scale upper ocean temperature variability, Prog. Oceanogr. 36 (1995) 169–217.

[16] White W.B., Meyers G., Hasunma K., Space/time scales of an oceanographic network in the mid-latitude North Pacific, J. Geophys. Res. 87 (1982) 1979–1989.

[17] Woodruff S.D., Slutz R.J., Jenne R.L., Steuer P.M., A comprehensive ocean atmosphere data set (COADS), Bull. Amer. Meteorol. Soc. 68 (1987) 1239–1250.

[18] WMO, World Ocean Circulation Experiment, Implementation plan: Detailed requirements, vol. 1, WMO/TD-No. 242, World Meteorological Organization, 1986, Geneva.