Quality Control and Interpolations of WOCE-TOGA Drifter Data

DONALD V. HANSEN

*NOAA/Atlantic Oceanographic and Meteorological Laboratory and Cooperative Institute
for Marine and Atmospheric Studies, University of Miami, Miami, Florida*

PIERRE-MARIE POULAIN

*Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida,
and SACLANT Undersea Research Centre, La Spezia, Italy*

ABSTRACT

Satellite-tracked drifting buoy data are being collected by numerous investigators and agencies in several countries for the World Ocean Circulation Experiment–Tropical Oceans Global Atmosphere (WOCE–TOGA) Surface Velocity Program. By the end of the century, and thereafter, this global dataset will provide the best available climatology and chronology of the surface currents of the World Ocean. To expedite completion of research quality datasets for archival and dissemination, a data acquisition activity is being conducted at NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML), Miami, Florida. At AOML, data from drifting buoys of cooperating operators are quality controlled and optimally interpolated to uniform 6-h interval trajectories for archival at the Marine Environmental Data Service (Canada). This report describes in detail the procedures used in preparing these data for the benefit of second- or third-party users, or future buoy operators who may wish to process data in a consistent way. Particular attention is given to provide quantitative estimates for uncertainty of interpolation.

1. Introduction

A global dataset of ocean surface currents and temperature is being collected for the World Ocean Circulation Experiment–Tropical Oceans Global Atmosphere (WOCE–TOGA) Surface Velocity Program (SVP) using the ARGOS system to collect data and find locations of free-drifting surface buoys. Standards for Lagrangian performance and designs to meet these standards have been promulgated through the WOCE–TOGA planning process. Regional buoy operations have been managed by numerous independent investigators; consequently, several observing schedules have been used, and there have been no generally agreed upon standards for processing such Lagrangian data. To expedite development of a uniform and timely research quality global dataset of ocean surface currents and temperature, a data acquisition center for the SVP was established at the NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML). This report is published to document the procedures used for the benefit of the many expected users of the global dataset, and to present ideas and statistical results that

may be of value to others who may wish to treat these or future drifter data differently to achieve their purposes.

Operation of an expansive array of drifting buoys began in the tropical Pacific Ocean in 1988. General overviews of the oceanographic results based on the subset of these data from the tropical Pacific appear in Reverdin et al. (1994) and P. Niiler et al. (1996; manuscript submitted to *J. Phys. Oceanogr.*). Subsets of these data have been used also in numerous special investigations (cf. Hansen and Maul 1991; Maul et al. 1992; Poulain et al. 1992; Poulain 1993). Drifting buoy releases were subsequently extended to the North and South Pacific, Atlantic, and Indian Oceans beginning in 1991.

The primary datasets consist of two variables: time series of buoy locations and sea surface temperature (SST) measurements from hull-mounted thermistors. All buoys reported in this archive were initially attached to drogues or sea anchors centered at 15-m depths to secure adequate water-following behavior and carried one of a variety of sensors for detecting and reporting the presence or loss of the subsurface drogue. Other measurements, such as barometric pressure, salinity, or chlorophyll, made on a very limited or experimental basis, are included in this archive, but at present only in the form of unedited sensor values as transmitted from the buoys and translated by Service ARGOS.

Corresponding author address: Dr. Donald V. Hansen, Environmental Research Laboratory, NOAA/AOML, 4301 Rickenbacker Causeway, Miami, FL 33149-1097.

Operators of SVP buoys are encouraged to have data from their buoys disseminated via the Global Telecommunications System (GTS) by Service ARGOS for use in operational SST analyses. These GTS transmissions are collected and placed also in the international drifting buoy data archive maintained by the Canadian Marine Environmental Data Service (MEDS). Due to insufficient knowledge or interest and sometimes technical or proprietary considerations, many SVP operators have not had data from their buoys disseminated via the GTS. SVP buoy data that were disseminated via the GTS are mixed with data from other kinds of drifting buoys, mostly not designed for measurement of current. Also, the SVP data transmitted via the GTS do not include information about presence or loss of the subsurface drogues because the GTS data format has no provision for it, and drogue sensor data often cannot be interpreted on a day-to-day basis anyway. In sum, the GTS-derived drifting buoy data available from the MEDS is incomplete and of mixed and unknown quality in respect to ocean currents.

AOML receives complete series of position data computed from Doppler measurements directly from Service ARGOS or, rarely, from the buoy operator. Position and temperature values that violate rather simple but carefully applied continuity criteria to be described are then edited from the series. These edited data are the best available in terms of both location and temperature, and therefore are most appropriate for uses such as pointwise matchups to satellite SST measurements. They are irregularly distributed in time, however, and therefore are generally ill conditioned for many kinds of analysis or even for satisfactory display. They are, therefore, interpolated to uniform 6-h intervals using an optimum interpolation procedure. The method used is that of kriging, which is commonly used for two- and three-dimensional analyses. In our application, latitude, longitude, and SST are treated as separate one-dimensional time series. All of these datasets (complete raw data, edited data, and interpolated data) are forwarded to the MEDS for archival and distribution to requestors.

Initial processing is usually completed within a month after acquisition of the data so as to be available to managers of buoy operations. The initial processing includes generation of the edited dataset and a tentative set of interpolations and uncertainty estimates. At 6-month intervals, a final processing is completed and data are forwarded to the MEDS. This 6-month update schedule is the reason for breaking the data into 6-month "batches" in the procedure described in sequel.

Our editing procedures are described in section 2. Section 3 provides a detailed description of the optimal interpolation algorithms, and our treatment of its keystone, the statistical structure function, is covered in section 4. Section 5 provides some explanation and perspectives on our procedures and use of the resulting datasets.

2. Data editing

a. Location data

Drifter locations given by the Service ARGOS tracking system contain errors arising from instrumental noise (mainly oscillator stability), buoy–satellite orbit geometry, number and distribution of messages received during a particular orbital pass, and inaccuracies in orbit and time coding. Service ARGOS provides probabilistic quality indices, called location classes, for all locations determined and offers users options for receiving only data meeting specified probable accuracies. Being probabilistic, these indices do not preclude the occurrences of occasional large errors, especially with the dominant observing schedule used in TOGA and WOCE, that is, transmissions made during one day of three (Hansen and Herman 1989). Rather than use these indices, we have elected to apply editing and interpolation procedures based on the full sequence of data values.

We hereafter present a technique for removing the occasional bad drifter locations from the raw ARGOS data. The method is based upon speed between consecutive locations. For simplicity, the method description and examples are restricted to positions in one dimension. The first step in processing either location or SST data is to arrange the data in temporal sequence. Suppose that we have a time series of positions $x_i = x(t_i)$ given at nonuniformly distributed times t_i. The mean velocity between two consecutive points is

$$U = \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} u(t)dt$$

$$= \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} \partial_t x(t)dt = \frac{(x_{i+1} - x_i)}{(t_{i+1} - t_i)} \quad (2.1)$$

and is, therefore, readily obtained from the discrete positions and times. A maximum value above which velocities are considered bad is determined with the help of our knowledge of the circulation characteristics in the study area and with observed velocity histograms (for the individual drifters and/or for all the drifters). Once a mean velocity between a pair of points is bad, we have to decide which point in the pair is the bad location. This problem is not trivial, due to the diversity of time intervals between sequential positions.

The simplest method progresses forward in time: for a bad velocity, the second point of the pair is removed and the velocity is recomputed using the next point in sequence. If the velocity again is bad, that point is removed, and so on until the velocity is below the maximum allowed. This simple editing technique is not always efficient. Because the technique is progressing forward, it can remove many points, although only one point is seriously in error. To avoid the above problem, we employ the following procedure using sequential positions in both forward and backward time directions.

Going forward, the initial point is verified as good and so flagged. If the required velocity to the second point exceeds a specified criterion, the second point is flagged as "bad," and the speed from the initial point to the sequential points is computed until another "good" point is found and flagged. This good point is taken as a new initial point for subsequent speed calculation, and so on through the data sequence for an individual buoy trajectory. The procedure is then repeated in the reverse time direction. All points flagged as good in both directions are given a global good flag, and any that were flagged as bad in either direction are assigned a global bad flag. For each group of bad global flags, the global flags are updated with flags from whichever of the forward and backward procedures contains the largest number of good flags, and all points with bad global flags are removed. Data are initially edited using a default speed criterion of 5 kt (257 cm s^{-1}), and results in the form of trajectories and speed histograms are checked. If spikes or other improbable features are found, the editing is repeated using a speed criterion of 3 kt (154 cm s^{-1}), and the results checked again. Any residual problems are handled by reediting gappy data strings in shorter segments or subjective judgment. Examples of the functioning of this logic are shown in Fig. 1.

This procedure is simply a method of using both forward and backward differencing of the position data together with the time interval information to minimize the rejection of valid data and acceptance of faulty data that can otherwise happen when a bad position following a long data interval is followed in turn by several valid data at shorter intervals, for example. The bottom panel of Fig. 1 suggests a counterexample, but this circumstance is most likely to occur in intervals of frequent observations, where loss of some data is least injurious, and its importance is further limited by use of a conservative velocity criterion.

b. SST data

Each satellite pass that provides a buoy location through the ARGOS system also yields several transmissions of sensor (SST) data. This larger number of data and a narrower range of physically meaningful values makes SST data easier to edit than location data. First, the median value of the several data transmissions of each satellite pass is computed and assigned the time of the location. Then values outside an acceptable SST range are deleted. For the tropical Pacific, we use the range 15°–35°C.

Remaining deviant SST values are removed by applying a temperature change criterion relative to the recent temperatures experienced by the buoy. The first valid temperature observation is determined subjectively and is labelled m_i. The subsequent values are then sequentially tested using
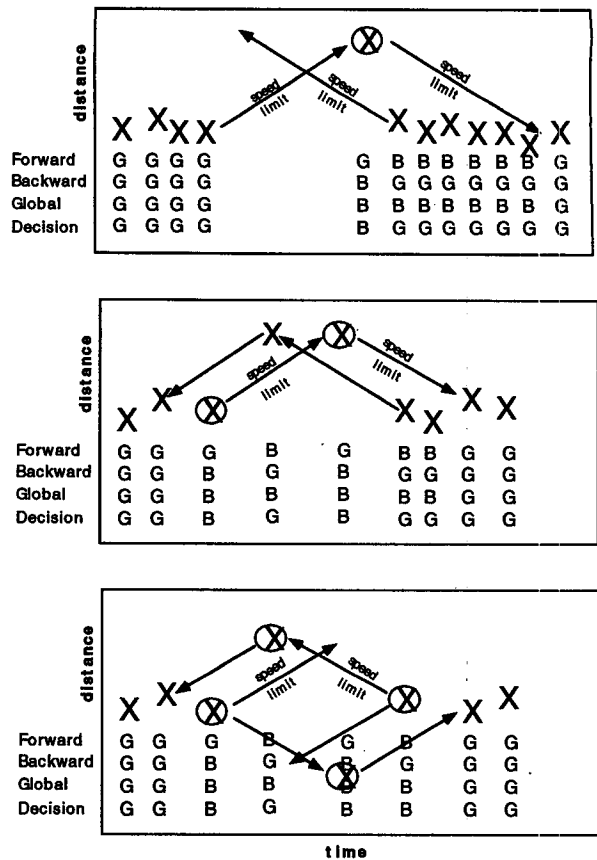

FIG. 1. Three synthetic illustrations of position editing logic. Upper panel: nf = 1, nb = 6, number of points rejected = 1; center panel: nf = 2, nb = 3, number of points rejected is 2; lower panel: nf = 2, nb = 2, number of points rejected is 4.

$$\mathrm{SST}_{i+1} - m_i | < \Delta, \qquad (2.2)$$

while m_i is updated by

$$m_{i+1} = C\mathrm{SST}_{i+1} + (1 - C)m_i, \qquad (2.3)$$

with $0 \leqslant C \leqslant 1$. SST values that fail (2.2) are deleted, and m_i is not updated. This procedure can be seen to be a detector of SST "spikes" that exceed a running weighted mean, in which the most recent values are most heavily weighted by an amount Δ or more. That is, an accepted SST value SST_{i-n} that occurred n values earlier is weighted as $(1 - C)^{n-1}$ relative to the most recent accepted value. To minimize improper data rejection when strong temperature fronts are crossed or large gaps occur in the data, the test is run forward and backward, and only data that fail the test in both directions are rejected. After some experimentation with data from the tropical Pacific, we selected the criteria $C = 1/4$, $\Delta = 0.5$ K. However, if more than 5% of the data from a particular buoy are rejected, Δ is increased by 0.1 K, at most twice, to a maximum of 0.7 K. If more than 5% of data still are rejected, the record is

subjectively examined for cause. Such results can be caused by multiple crossings of strong SST fronts. The data may be valid, and potentially valuable for evaluation of heat transports, but difficult to edit tightly enough with the automated procedure.

3. Interpolation of data

We employ a one-dimensional revision of a procedure known as kriging that is mostly used for two- and three-dimensional analyses in earth sciences and mining engineering (Cressie 1991). It is applicable to interpolation of data that are irregularly distributed, for which there is no useful norm or first guess, and for which the autocorrelation may not exist. We apply it to one-dimensional time series data for latitude, longitude, and SST. The method is similar in most respects to the form of optimum interpolation that was introduced to atmospheric and ocean scientists by Gandin (1963), but has some significant differences in its usual implementation. Our approach is similar to that used in Hansen and Herman (1989) except that treatments of measurement errors and structure functions are improved.

Interpolated values are estimated as linear combinations of a number of observations neighboring in time. We take each observation x_i to be a sum of the true value $\hat{x}_i$ and a measurement error that has a mean value of zero and is temporally uncorrelated with either itself or the true values. Thus,

$$\left.\begin{aligned} x_i &= \hat{x}_i + e_i \\ \langle x_i \rangle - \langle \hat{x}_i \rangle &= \langle e_i \rangle = 0 \\ \langle e_i e_i \rangle &= \langle e^2 \rangle \\ \langle e_i e_j \rangle &= 0, \quad \text{if} \quad i \neq j \end{aligned}\right\}, \quad (3.1)$$

where angle brackets are used to denote an average. Interpolated values are determined as best linear combinations of n neighboring observed values, that is,

$$\hat{x}_0^* = \sum_{i=1}^{n} w_i x_i, \quad (3.2)$$

in which $\hat{x}_0^*$ denotes an estimated value (without measurement error), the x_i are observed values, and the w_i are the set of best weights. "Best" can be qualified in many ways. Here we choose it to minimize the mean-square difference between the true values at the interpolation points and their estimates from (3.2)

$$\sigma_k^2 = \langle (\hat{x}_0 - \hat{x}_0^*)^2 \rangle = \text{minimum}. \quad (3.3)$$

Methods of optimum interpolation, familiar to oceanographers and meteorologists, use an expression obtained from equations corresponding to (3.2) and (3.3) containing the autocorrelation of the interpolation variable to determine the weights w_i. Because our interpolation variable is location, its autocorrelation function

tion cannot in general be expected to exist, for the same reason that it does not exist for a simple Brownian motion, that is, its variance increases without limit, or in practical terms, to the dimension of the ocean basin, which is of little help locally. Rather, we must use the structure function

$$S_{ij} = \frac{1}{2} \langle (x_i - x_j)^2 \rangle = \frac{1}{2} \langle [x(t_i) - x(t_j)]^2 \rangle, \quad (3.4)$$

which, in addition to being of more general existence, is more resistant to data defects. It is immediately obvious that it is independent of errors in estimation of the mean value of x, for example. To obtain an expression for the optimal weights in (3.2) from (3.3), in terms of the structure function, it is necessary to apply an additional constraint, that the n weights sum to unity,

$$\sum_{i=1}^{n} w_i - 1 = 0. \quad (3.5)$$

This constraint has the effect of assuring that the estimated values are unbiased in the mean, that is,

$$\langle (\hat{x}_0 - \hat{x}_0^*) \rangle = 0, \quad (3.6)$$

which is a useful property in itself inasmuch as we have not discovered any useful analog of the norm or first-guess fields commonly used in oceanographic and atmospheric applications of optimum interpolation. In such applications, where (3.5) is not used, the interpolations obtained from (3.2) are biased toward zero in regions of sparse data, so that the climatology, first guess, or other norm becomes the dominant element in the final analysis. With (3.5) as a constraint, the mean of the observations becomes the "default" value.

Substitution of (3.2) into (3.3) and use of (3.1) and (3.5) leads to

$$\sigma_k^2 = 2 \sum_{i=1}^{n} w_i \hat{S}_{0i} - \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \hat{S}_{ij} + \sum_{i=1}^{n} w_i^2 \langle e^2 \rangle \quad (3.7)$$

as the expression to be minimized with respect to w_i, subject to the constraint of (3.5); $\hat{S}$ denotes the structure function for error-free data and $\hat{S}_{0i}$ that for the interpolation point (x_0) relative to the observations (x_i). Adding the constraint with a Lagrange multiplier of 2λ, differentiating with respect to the w_i and λ, and equating the result to zero leads to

$$\sum_{i=1}^{n} w_i \hat{S}_{ij} - w_j \langle e^2 \rangle + \lambda = \hat{S}_{0j}, \quad (3.8)$$

as the set of equations for determining optimum weights. Multiplication of (3.8) by $\sum_{i=1}^{n} w_i (=1)$ to eliminate the double summation from (3.7) provides a simpler expression for the kriging variance, σ_k^2:

$$\sigma_k^2 = \sum_{i=1}^{n} w_i \hat{S}_{0i} + \lambda. \quad (3.9)$$

Following some experimentation, we settled on n = 10, five observations preceding and following each interpolation point when available, as a suitable compromise between computation speed and quality of result. When data are few, as at the beginning or end of a buoy life, or where there are long data gaps, interpolations are also done with as few as a single observation on either side of the interpolation time. Both the interpolated values and their associated estimation errors (σ_k) are contained in the archive data.

Illustrative 16-day time series of edited and interpolated latitude and longitude data are shown in Fig. 2. Also indicated, in exaggerated form, are the σ_k expressing the varying uncertainty of the interpolation. The manner in which the σ_k are modulated by the density and temporal distribution of observations is evident. Figure 3 shows the buoy trajectory generated from these data, also with σ_k values. It should be kept in mind that the σ_k are statistical estimates equal to one standard deviation, so they will be exceeded in about one-third of the interpolation. In section 4 we describe our procedure to make the σ_k as quantitative as possible, but they may be either conservative or optimistic for individual points or, due to temporal nonstationarity and spatial nonhomogeneity of the ocean, for major segments of, or even entire trajectories.

4. Specification of structure functions

a. Location data

It is evident that the structure function plays a key role in the interpolation. It must be known from prior data or determined from the observations themselves. Equations (3.8) and (3.9) require the structure function for error-free data but the observations contain measurement errors that, for our edited data, are small relative to the geophysical variance, but which we chose to include nonetheless. Use of (3.1) in (3.4) gives the relationship

$$S_{ij} = \hat{S}_{ij} + \langle e^2 \rangle - \langle e_i e_j \rangle \qquad (4.1)$$

between the structure function for observations and that for error-free data. Hence, for $i \neq j$, $\hat{S}_{ij} = S_{ij} - \langle e^2 \rangle$, and for $i = j$, $\hat{S}_{ij} = S_{ij} = 0$ (i.e., S_{ij} is discontinuous at the origin). The measurement error $\langle e^2 \rangle$ was estimated from historical data from drifting buoys operated in the tropical Pacific Ocean (Hansen and Herman 1989; Bitterman and Hansen 1993).

Empirical values of the structure functions were computed from observations representative of data from the tropical Pacific to be interpolated. Data from 1988 through 1993 were sorted into 11 6-month (January–June, July–December) batches for three zonal bands defined by latitudes 23.5°N, 7°N, 3°S, and 23.5°S. This regionalization was motivated by evaluations of the distribution of kinetic energy density such as that shown by P. Niiler et al. (1996; manuscript sub-
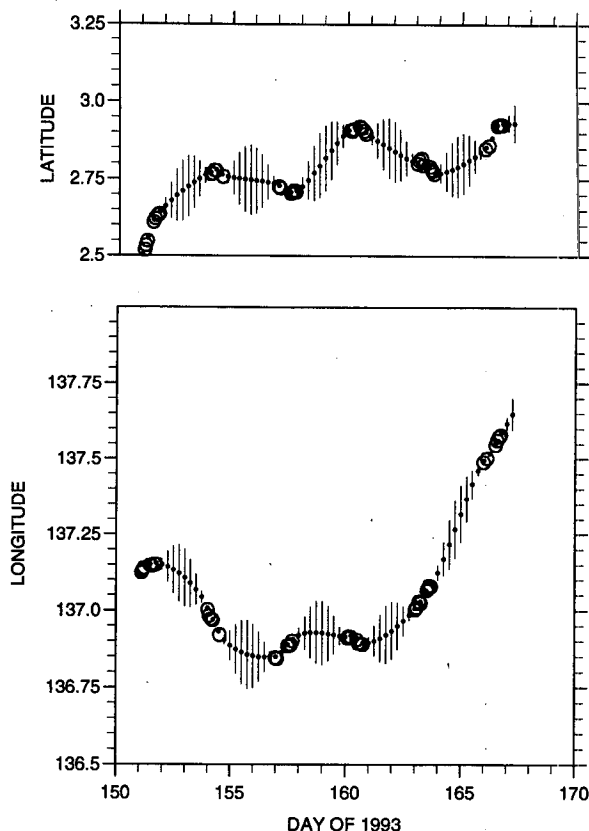

FIG. 2. Illustrative time series of edited (circles) and interpolated (dots) latitude and longitude coordinates for drifting buoy 2898. Vertical bars through interpolated points show magnitude of $20\sigma_k$.

mitted to *J. Phys. Oceanogr.*), and an expectation that the structure function will reflect the kinetic energy density, especially at the shorter time separation of importance here. Squared differences of the observations were binned at quarter-day lag intervals and averaged. The locus of such empirical values is typically concave upward for lags to several days, becoming more linear for lags of a few weeks, and becoming very irregular due to the smaller number of data pairs available for averaging at long lags. Results shown by Hansen and Herman (1989) are illustrative.

To provide structure function values at all possible lags needed for the interpolations, the empirical values must be modeled by a conditionally negative definite function (Cressie 1991). We used a fractional Brownian motion model (Mandelbrot and Van Ness 1968),

$$\hat{S} = \alpha \tau^\beta, \qquad (4.2)$$

in which τ denotes time (nondimensionalized by 1 day) and α and β are parameters to be determined from the empirical data. This model is well known in the literature on kriging; it is conditionally negative definite for $0 \leq \beta < 2$, and has been found useful for agricultural and hydrologic applications (Cressie
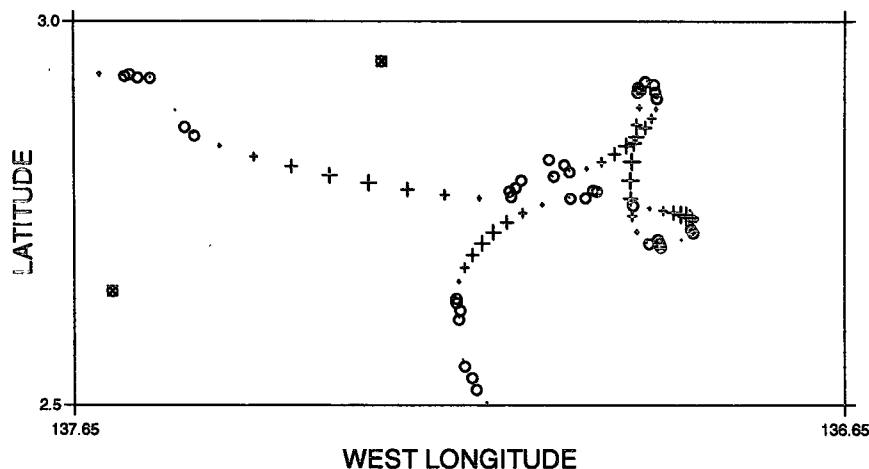
FIG. 3. Illustrative example of location editing and interpolation results. A "O" denotes location data as received from Service ARGOS, actual position is at center of circle, diameter does not imply precision; a "×" denotes a rejected point; and "+" denotes interpolated locations at the center, with arms equal to $\pm 2\sigma_k$.

1991). For drifting buoy movements, the physically meaningful range is $1 \le \beta < 2$. At $\beta = 1$, the model is that for Brownian motion, or pure diffusion (Einstein 1905). The limit, $\beta = 2$, is readily seen to be the model for rectilinear motion. For such motion, from (3.4),

$$(x_i - x_j) \rightarrow u(t_i - t_j) = u\tau,$$

in which u denotes the (constant) value of the velocity component, and hence $\hat{S} = u^2\tau^2$. This model is not usable for kriging because it is not conditionally negative-definite. That is of no concern here because drifting buoys do not show rectilinear motion for more than brief intervals. (Were it to occur, it would imply the possibility of exact, linear interpolation.)

The WOCE–TOGA SVP data from the tropical Pacific are fit by values of β near 1.6 for latitude data and 1.8 for longitude data, indicating properties intermediate between the extremes of an uncorrelated random walk (diffusion) and perfectly correlated motion (rectilinear advection), with advection being of greater influence in the zonal direction. We have not evaluated the full range of validity of this model for ocean surface particle displacements. It appears to be usefully accurate for timescales from a few hours to a few days, at least in tropical regions, and is a useful precept for statistical interpolation of quasi-Lagrangian buoy data. Hansen and Herman (1989) show some examples of least square fits of the fractional Brownian motion model to data for lags of up to six days. Although these fits to the empirical data appear to be excellent, twofold differences occurred between the kriging variances and the values obtained by cross validation, indicating that least square fits, even to seemingly regular data, do not necessarily provide the best results. In general, the kriging variances associated with the interpolation tend to

be more sensitive to changes in model parameters than are the interpolations themselves. To evaluate the range of lags over which the model should be tuned, we determined the distribution function of data intervals used for estimation of kriging variances (3.9) for approximately 10^6 interpolations in the tropical Pacific. The median time separation used was about 1.5 days, the mean about 2 days, and the 95th percentile was about 5.4 days. Hence, we conclude that the critical range for accuracy of the structure functions is zero to about 1 week.

The smoothly varying empirical structure function values shown by Hansen and Herman (1989) were derived from data from buoys in continuous transmission, that is, usually five or six locations were received each day. In the TOGA and WOCE SVP programs, however, the principal observing schedule was for buoy transmissions during one day of three. Each day of transmissions is expected to provide several locations during that day. Such a sampling schedule, if uniformly implemented, would yield no empirical structure function values at all for lags of one to two days, four to five days, etc. For various reasons, individual investigators have used other transmission schedules for all or a part of the operating life of their buoys. The composite dataset, therefore, also contains varying amounts of data from buoys in continuous transmissions, transmissions during 1 day of 2, and during 8 hours of 24. Also, buoys programmed for transmissions during one day of three occasionally shifted the phase of their transmission duty cycle. The admixture of observing schedules provides large numbers of data pairs for computing structure function near lags of 3, 6, 9, etc., days when many buoys contribute data, but smaller numbers of data at the intervening lags. The structure
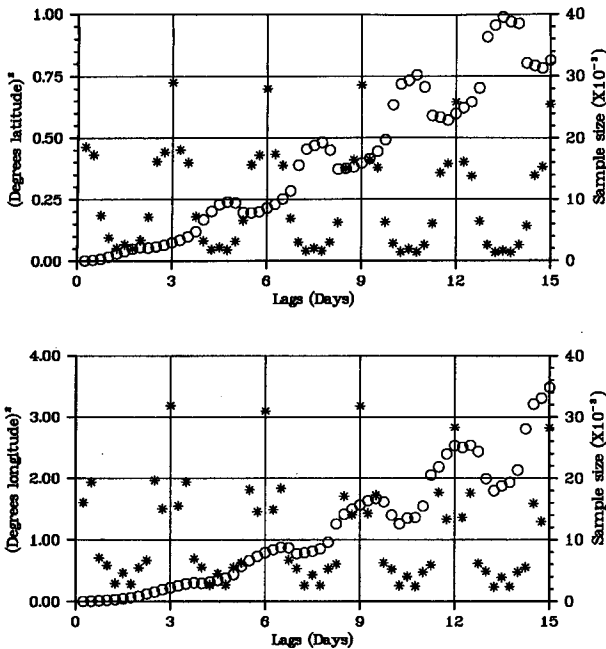
FIG. 4. Empirical structure function values for two 6-month batches of drifting buoy data. Circles denote empirical structure function values, asterisks denote number of difference-pairs used. Upper panel: data for latitude, north tropical region, first half of 1992; lower panel: data for longitude, equatorial region, first half of 1993.

function values computed at these intervening lags are subject to the usual perils of averages from small samples and may be biased toward the particular time and region where an individual investigator chose to use, for example, continuous transmissions for 2 or 3 months. In contrast to the smooth loci of empirical values of structure functions shown by Hansen and Herman (1989), those derived from the composite WOCE–TOGA datasets often shown large variations at multiples of 3-day lag intervals that reflects the varying number of data pairs. Figure 4 shows examples for which the structure function has maxima (minima) near 3, 6, 9, etc., days. A fit of the model to such data by a simple least squares procedure is inappropriate. Quantification of the model might be improved by a weighted least squares procedure, or simply by deleting from the fit structure function values derived from too few data. These possibilities carry undesirable degrees of arbitrariness in specifying a weighting scheme or deciding just what number of data is too small. Considering that the range of validity of the structure function model need not exceed about a week, that the two best determined values are at 3 and 6 days, and that the results of Hansen and Herman (1989) show that structure functions computed from uniform and higher frequency data are smoothly varying, we decided upon a two-step procedure for quantification of the structure function model using a first-guess based on a deter-

ministic fit of the model to the empirical values at 3- and 6-day lags. Thus, for (4.2) we use

$$
\left.
\begin{aligned}
\beta &= \frac{1}{\ln 2} \ln\left(\frac{\hat{S}_6}{\hat{S}_3}\right), \\
\alpha &= \frac{\hat{S}_3}{3^\beta},
\end{aligned}
\right\}
\tag{4.3}
$$

in which $\hat{S}_3$, $\hat{S}_6$ are the empirical structure function values [adjusted for measurement error as per (4.1)] at 3- and 6-day lags.

The median values of the 11 determinations of α and β for each of the three regions are listed in Table 1. It is evident that the exponent β is systematically larger for longitudinal than for latitudinal displacement data, presumably reflecting the stronger mean zonal advection. The amplitude coefficients α are about twice as large for longitude data as for latitude data and also strongly reflect the distribution of kinetic energy density (P. Niiler et al. 1996; manuscript submitted to *J. Phys. Oceanogr.*) that motivated regionalizations of the structure function model. The extremes of temporal (batch) variations from the median values given in Table 1 never exceeded 7% for β, but varied from nearly 40% to nearly 100% for α.

The model coefficient α is a simple multiplier that nearly factors from (3.7) because measurement error in our data is small. The interpolations, therefore, have small sensitivity to variations of α, but must be expected to vary with β. The associated values of σ_k^2, however, are nearly proportional to α as well as sensitive to β. Our objective is to obtain quantitative estimates of σ_k^2 as well as optimum interpolation. The accuracy of the σ_k^2 values can be evaluated batchwise by cross validation. Observations are sequentially deleted and their value estimated from neighboring observations using the kriging algorithm with the structure function model under test. The average mean square difference $\langle \delta^2 \rangle$ between these estimates and the actual observations can then be compared with the average value of σ_k^2 for the interpolation. If the σ_k^2 estimates are accurate, $\langle \sigma_k^2 \rangle$ should be close to $\langle \delta^2 \rangle - \langle e^2 \rangle$. For the first guess at α and β, we use median values from 11 6-month data batches from the tropical Pacific (Table 1). For final processing, the cross-validation

TABLE 1. Generic or first-guess values of displacement structure function parameters determined from 6.5 years of drifter data from the tropical Pacific Ocean.

	Latitude		Longitude	
Region	$10^2\alpha(°\phi)^2$	β	$10^2\alpha(°\lambda)^2$	β
North tropical	1.152	1.604	1.912	1.846
Equatorial	1.896	1.538	5.151	1.854
South tropical	0.596	1.673	1.129	1.861

test is run on each batch using the first-guess values of α and β, and α is adjusted by

$$\alpha' = \frac{\langle \delta^2 \rangle - \langle e^2 \rangle}{\langle \sigma_k^2 \rangle} \alpha, \qquad (4.4)$$

which assures that $\langle \sigma_k^2 \rangle \Rightarrow \langle \delta^2 \rangle - \langle e^2 \rangle$, batchwise, in the final processing for archival.

b. SST data

WOCE–TOGA SVP buoys are designed for nominal accuracy of 0.1 K in measurement of SST, which Bitterman and Hansen (1993) show seems to be mostly realized. In some data batches, we have found S (1 day) < 0.01 K^2, indicating a smaller than expected measurement error. Hence, for SST, empirical values of the structure function are evaluated for each data batch, and $\langle e^2 \rangle$ is defined as the lesser of 10^{-2} K^2 or 0.5 S (1 day).

Interpolation of SST measurements could be done in terms of departures from a norm, such as Reynolds (1982) climatology of SST, obviating the need for the constraint (3.5) on optimization of the weights, which in principle should make them "more optimal." We elected not to do this in order to keep the interpolations as close as possible to the measurements, albeit with larger kriging variance, even in the circumstance of low data density in the presence of large thermal anomalies, such as El Niño. Given the large scale of patterns in both SST climatology and El Niño variations, the short-term structure functions for the departure data are expected to be very similar to those for the full measurements. Kriging variances for SST are sufficiently small as not to be a major concern in any case (Hansen and Herman 1989).

A substantial part of the zero to several day variance of SST measurements is associated with the diurnal cycle (Hansen and Herman 1989; Bitterman and Hansen 1993) and can be included in the model. Thus, for SST we use

$$S = \alpha \tau^\beta + \gamma(1 - \cos 2\pi\tau). \qquad (4.5)$$

Unlike the model for buoy displacements, there seems to be no a priori reason to expect that β should not be less than unity. Usually a value near unity has been found for β from the drifting buoy temperature data, a result presumed by Hansen and Herman (1989) and Bitterman and Hansen (1993).

Because the factor $(1 - \cos 2\pi\tau)$ is zero at 3- and 6-day lags, (4.3) can be used to evaluate α, β for the structure function model for SST, just as for buoy displacements. To avoid the problems associated with widely varying numbers of data pairs, we estimate γ from the empirical structure function value at 0.5-day lag,

$$\gamma = 0.5[\hat{S}(0.5) - \alpha(0.5)^\beta]. \qquad (4.6)$$

Finally, a different regionalization was defined for the structure function model parameters for SST than

for displacements. We evaluated the parameters for six regions spanning the tropical Pacific, again using 5.5 years of buoy measurements in 6-month batches. Parameters derived from measurements made west of 130°W are nearly equal, and can be treated in common. Values obtained with data from opposite sides of the equator, east of 130°W, were sufficiently different from those for the west–central Pacific, and from each other, as to merit separate treatments. The resulting first-guess values are listed in Table 2. The magnitude of the rms diurnal cycle, and the nearly linear rate of increase of the several-day variance are common to all regions. The magnitude of the several-day variances varies in the approximate ratio 1:2:3, reflecting the low variability of SST in the west–central tropical Pacific, and the large SST variability associated with the equatorial front (Philander et al. 1985) and coastal processes along the coast of Central America. The moderate value of α for the southeast region appears to be inconsistent with the SST-variance map shown by P. Niiler et al. (1996; manuscript submitted to J. Phys. Oceanogr.), until it is understood that much of this variance occurs on annual and interannual timescales and thus does not affect the structure function or the interpolation on timescales of a few days.

5. Discussion

Two relatively different situations in interpolating the WOCE–TOGA SVP data can be distinguished. One is that of interpolating data during days when a buoy is transmitting so that there are typically several observations within a day of the interpolation time. The other is that of interpolating across the 2-day transmission gaps, where at least half of the interpolation times have no observations within 1 day. Hansen and Herman (1989, Fig. 3) show the temporal modulation of σ_k^2 that results from this varying pattern of data density. The cross-validation method that we use to adjust $\langle \sigma_k^2 \rangle$ for equality with the mean square misfit of estimates to observations is more relevant to the transmission days than to the gaps because the cross-validation data necessarily come from days with transmissions and typically have near neighbors. Because the structure function model contains two parameters, we sought a method by which they could be adjusted jointly to satisfy $\langle \sigma_k^2 \rangle = \langle \delta^2 \rangle - \langle e^2 \rangle$ for two different densities of data; $\langle \sigma_k^2 \rangle = \langle \delta^2 \rangle - \langle e^2 \rangle$ for a given dataset can be

TABLE 2. Generic or first-guess values of SST structure function values determined from 6.5 years of drifter data from the tropical Pacific Ocean.

Region	$10^2\alpha$	β	$10^2\gamma$
West-central	0.65 K^2	1.0	0.7 K^2
Northeast	1.70 K^2	1.0	0.7 K^2
Southeast	1.20 K^2	1.0	0.7 K^2

thought of as defining a curve in the (α, β) plane. If the density of this dataset is reduced by deleting observations, a different curve is defined. An intersection of these two curves would give (α, β) meeting our objective at both data densities, and hopefully be close at intervening densities as well. This hope foundered on the result that in about 10% of the cases we tested, the curves do not intersect within the range $1 \leq \beta < 2$, and for the other 90%, the intersections tend to be at very small angles, causing large batch-to-batch variations, especially in β. We therefore elected to use regionally generic values of β and to adjust only the more variable α by the batch cross-validation process. The extreme range of $\langle \sigma_k^2 \rangle (\langle \delta^2 \rangle - \langle e^2 \rangle)^{-1}$ obtained by this process for the full edited data batches was 0.90–1.02, and the median value was 0.99. We also tested the robustness of the structure functions to variation in temporal density of the measurements. The cross-validation procedure was applied also to the same data batches thinned by deleting all measurements during the day of the interpolation except the first and last, thus forcing the algorithm to use measurements more distant in time. For this thinned dataset, the fractional Brownian motion model using generic coefficients in Table 1 followed by adjustment using (4.4) yielded values of $\langle \sigma_k^2 \rangle (\langle \delta^2 \rangle - \langle e^2 \rangle)^{-1}$ with an extreme range of 0.41–1.24, and the median values listed in Table 3. Thus, the kriging variances that are archived with the interpolations are tuned to be accurate during times of plentiful measurements and to become conservative, more so for the more energetic variables, when measurements are sparse.

We experimented also with use of an exponential model that is familiar in the literature on kriging and a model,

$$\hat{S} = \alpha \tau^2 (\tau + 2\beta)^{-1}, \qquad (5.1)$$

which at short lags behaves similarly to the rational quadratic function (Cressie 1991), but which appears to be new to the literature. We have qualified it as conditionally negative-definite using the procedures described by Christakos (1984). It is attractive on physical grounds in that the parameters α, β can be identified with the diffusivity and integral timescale from the well-known results of Taylor (1921) for homogeneous turbulence. That is, for large lags, τ, $\hat{S} \to \alpha \tau$, so that α is equivalent to the diffusivity D as defined by Taylor. For small lags, $\hat{S} \to \alpha/\beta 2\tau^2$ and $\hat{S} \to 0.5 \langle v^2 \rangle \tau^2$, so that $\beta = k/\langle v^2 \rangle = T$, the integral timescale. In our application, however, this model proved less robust to variations in temporal density of measurements than was the fractional Brownian motion model. The median values of $\langle \sigma_k^2 \rangle (\langle \delta^2 \rangle - \langle e^2 \rangle)^{-1}$ were not notably inferior to those shown in Table 3, but the range increased to 0.97–2.49 for full-data cross validation, and 0.28–2.10 for thinned data. Fractional Brownian motion is said to be "self-similar," a form of invariance with respect to scale that makes properties determined at a

TABLE 3. Median values of kriging variance ratios from thinned data batches.

Region	Latitude	Longitude
North tropical	0.93	0.81
Equatorial	0.80	0.58
South tropical	0.86	0.79

particular scale applicable also at adjacent scales (Mandelbrot and Van Ness 1968). This appears to explain the relative robustness of kriging variances obtained with this model to variations of data density, and adds to the growing body of evidence that ideas about fractals have relevance to drifter data.

The model, (5.1), had another interesting defect in that values of the parameters α, β proved to be quite sensitive to the length of lag interval used for fitting the model to data. In particular, the diffusivity and the integral timescale both increased with the range of the lag interval to which the model was fit. Sanderson and Booth (1991) show that this is characteristic of fractional Brownian motion for which an integral timescale is computed in the usual way by integration of the velocity autocorrelation function. The value of β implied by fractal dimensions of drifter data from the northeast Atlantic by Sanderson and Booth is 1.3–1.6 for scales of 5–100 km. Comparison of these values to Table 1 suggests that the northeast Atlantic is a relatively more diffusive regime than is the tropical Pacific, as one might expect from the more prominent appearance of eddies in higher latitudes generally. One obvious problem with use of (5.1) is that, lacking a useful norm, we elected to use the full displacement data rather than a "turbulence" component in our structure function and interpolation. It is mentioned here with the thought that it may be more useful in application to deviation data at such future time when the surface circulation climatology becomes sufficiently well described to provide a useful norm.

Data from WOCE–TOGA sometimes contain gaps of several days. Interpolations are made across gaps not exceeding one month. The associated large values of σ_k serve as a flag that these interpolations should be used judiciously. They can be used, for example, for computation of pseudo-Eulerian average velocities if the size of the averaging elements is sufficiently large relative to σ_k for locations. The velocity averaging property of Lagrangian data can be used effectively to control errors so long as it is assured that a drifter continuously occupies the averaging domain. Applications requiring more precise velocity data over shorter intervals can be satisfied by computing velocities by differencing locations with low σ_k values, which mostly occur at 3-day intervals.

Most of the particulars of structure function models given in this report are based on observations made

over a period of years in the tropical Pacific. These results will be used as a default strategy for other tropical oceans until sufficient observations have been made to confirm or change them. For higher latitudes, it seems desirable to model effects of inertial oscillations as well as possible. A model similar to that used for the diurnal variation of SST, but with frequency a function of latitude, seems a possibility, but considerable new observations are needed for quantification.

REFERENCES

Bitterman, D. B., and D. V. Hansen, 1993: Evaluation of sea surface temperature measurements from drifting buoys. *J. Atmos. Oceanic Technol.,* **10,** 88–96.

Christakos, G., 1984: On the problem of permissible covariance and variogram models. *Water Resour. Res.,* **20,** 251–265.

Cressie, N. A. C., 1991: *Statistics for Spatial Data.* John Wiley and Sons Inc., 900 pp.

Einstein, A., 1905: Über die von der molecularkinetischen Teorie der Wärme geforderte Bewegung von in ruhenden Flussigkeiten suspendierten Teilchen. *Ann. Phys.,* **17,** 549.

Gandin, L. S., 1963: *Objective Analyses of Meteorological Fields* (in Russian). Gidrometeorologicheskoe Izdatel'stuo, 242 pp.

Hansen, D. V., and A. Herman, 1989: Temporal sampling requirements for surface drifting buoys in the tropical Pacific. *J. Atmos. Oceanic Technol.,* **6,** 599–607.

——, and G. A. Maul, 1991: Anticyclonic current rings in the eastern tropical Pacific Ocean. *J. Geophys. Res.,* **96**(C4), 6965–6979.

Mandelbrot, B. B., and J. W. Van Ness, 1968: Fractional Brownian motions, fractional noises, and applications. *SIAM Rev.,* **10,** 422–437.

Maul, G. A., D. V. Hansen, and N. J. Bravo, 1992: A note on sea level variability at Clipperton Island from GEOSAT and in-situ observations. *Sea Level Changes: Determination and Effects, Geophys. Monogr.,* **69**(11), IUGG, 145–154.

Philander, G., D. Halpern, D. V. Hansen, R. Legeckis, L. Miller, C. Paul, R. Watts, R. Weisberg, and M. Wimbush, 1985: Long waves in the equatorial Pacific Ocean. *Eos, Trans. Amer. Geophys. Union,* **66,** 154.

Poulain, P.-M., 1993: Estimates of horizontal divergence and vertical velocity in the equatorial Pacific. *J. Phys. Oceanogr.,* **23,** 601–607.

——, D. Luther, and W. Patzert, 1992: Deriving inertial wave characteristics from surface drifter velocities: Frequency variability in the tropical Pacific. *J. Geophys. Res.,* **97**(C11), 17 947–17 959.

Reverdin, G., C. Frankignoul, E. Kestenare, and M. J. McPhaden, 1994: Seasonal variability in the surface currents of the equatorial Pacific. *J. Geophys. Res.,* **99**(C10), 20 323–20 344.

Reynolds, R. W., 1982: A monthly averaged climatology of sea surface temperature. Tech. Rep. NOAA NWS-31, Silver Spring, MD, 35 pp.

Sanderson, B. G., and D. A. Booth, 1991: The fractal dimensions of drifter trajectories and estimates of horizontal eddy diffusivity. *Tellus,* **43A,** 334–349.

Taylor, G. A., 1921: Diffusion by continuous movements. *Proc. London Math. Soc.,* **20,** 196–211.