Sifting out Erroneous Observations in COADS - The Trimming Problem

Klaus Wolter
Cooperative Institute for Research in Environmental Sciences
University of Colorado at Boulder, Boulder, CO 80309-0449

1. Background

The Comprehensive Ocean-Atmosphere Data Set (COADS; Slutz et al., 1985; Woodruff et al., 1987) is currently the most complete marine surface data set available. While a variety of formats is offered, monthly trimmed 2°x2° summaries are the most widely used data. Trimming refers to the process of eliminating statistical outliers from the data. Such outliers are currently defined with respect to sextiles based on all observations within a calendar month and during the decades within a trimming period (1950-79) for the presented results). For a given 2°x2° square and calendar month, upper/lower standard deviations (σ_u/σ_l) are defined as the difference between the 5th/1st sextile and the median (G) of all observations (Woodruff et al., 1987). Sextile were slightly modified in order to yield standard σ for normal distributions. If an observation falls above $G+3.5*\sigma_u$ or below $G-3.5*\sigma_l$, it is removed from the trimmed product. Thus, even asymmetric observational distributions are screened in a seemingly robust manner. Note that interim 1980-91 products have been trimmed using the 1950-79 limits.

Based on earlier studies (Wolter, 1989; Wolter et al., 1989), this paper briefly recapitulates some of the errors that have occurred using this trimming procedure, and discusses possible remedies. In principle, it can either happen that an extreme, but realistic observation is erroneously removed (statistical Type I error), or that an unrealistic, outlying observation is not removed (Type II error). Extreme but real Pacific SST data have indeed been trimmed repeatedly in COADS (Wolter et al., 1989). Overall, Type I errors have been ignored until recently, even though they could hamper our ability to study the variability of climate through the more than century-long COADS record. The reader is referred to Trenberth et al. (1992) for many of the error sources, such as errors due to instrumentation (changes) and due to inhomogenous sampling in space (shiptracks) and time (diurnal cycle, within-monthly changes due to the seasonal cycle). Modem telecommunicated (GTS) data introduce the additional problem of relatively frequent missing or otherwise altered (leading) digits, especially during the last decade. Conversion errors such as from Fahrenheit to Celsius, or from knots to m/s, as well as errors in reported position are further sources of confusion (Woodruff et al., 1992). The impact of these errors is aggravated in large parts of the World Ocean due to low sampling densities.

Tropical Pacific SST is particularly vulnerable to Type I trimming errors. This is due to a combination of factors: large year-to-year variability associated with the El Niño/Southern Oscillation phenomenon, high intermonthly persistence for extreme conditions, and relatively small within-monthly variability and spatial gradients away from the upwelling regions. As defined above, sextile-based trimming limits from 1950-79 do not admit the full extent of the strongest event of the century (1982/83; Wolter, 1989). In fact, the addition of the 1980’s to the base trimming period of 1950-79 would quite likely still not do it justice, 1982/83 being only a small fraction of the total data.

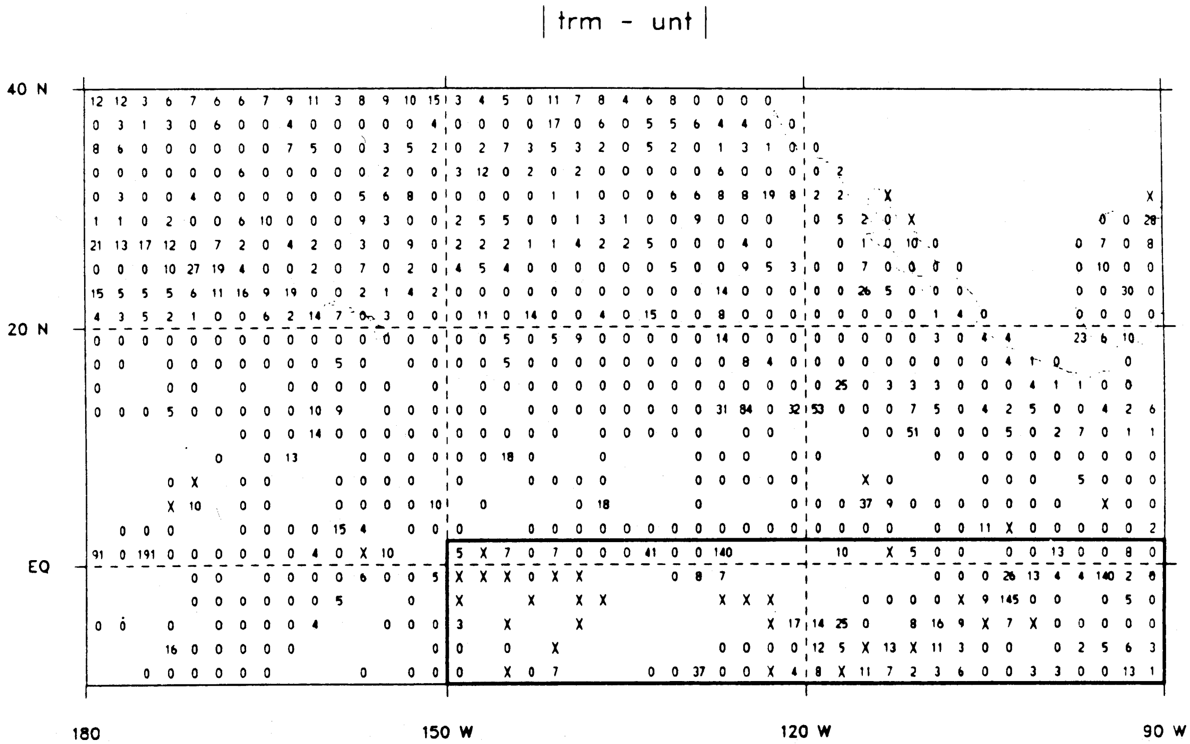

Figure 1. Absolute differences [in 1/10°C] between trimmed and untrimmed COADS SST in February 1983 for the eastern and subtropical Pacific. If the trimming process removed all data from a given 2°x2° square, it was marked by an 'X'. The chosen subdomain with particularly heavy losses due to trimming is indicated by a solid outline [2°N-10°S, 90°-150°W].

The extent of the problem is illustrated in Fig. 1 for February 1983 SST in the -eastern Pacific. The remainder of this paper will focus on the eastern equatorial Pacific region most affected by the trimming problems [2°N - 10°S, 90° - 150°W]. It coincides largely with the so- called ENSO Region 3 used for monitoring the state of this phenomenon, hence it will be referred to as ENSO 3. Effects of the trimming procedure on the analyzed 1982/83 SST anomalies are presented in more detail in section 2. Alternative trimming procedures for SST are discussed in section 3. The paper finishes up with one final alternative trimming procedure (section 4.) that appears to be the most promising general solution to this problem. Its algorithms are presently being tested, and will be considered for implementation in developing a separate set of newly trimmed monthly 2°x2° summaries for 1980-91 SST data.

2. Case-Study Eastern Tropical Pacific SST in 1982/82

Figure 2 shows the percentage of observations trimmed within ENSO 3 during 1982/83. In this region, February 1983 was the peak loss month of the trimming process. This occurred two months after December 1982 when the actual SST anomalies peaked (see below, Fig. 5). Since the current trimming limits derive from standardized departures, this reflects the delayed peak in standardized SST anomalies during the early spring of 1983.

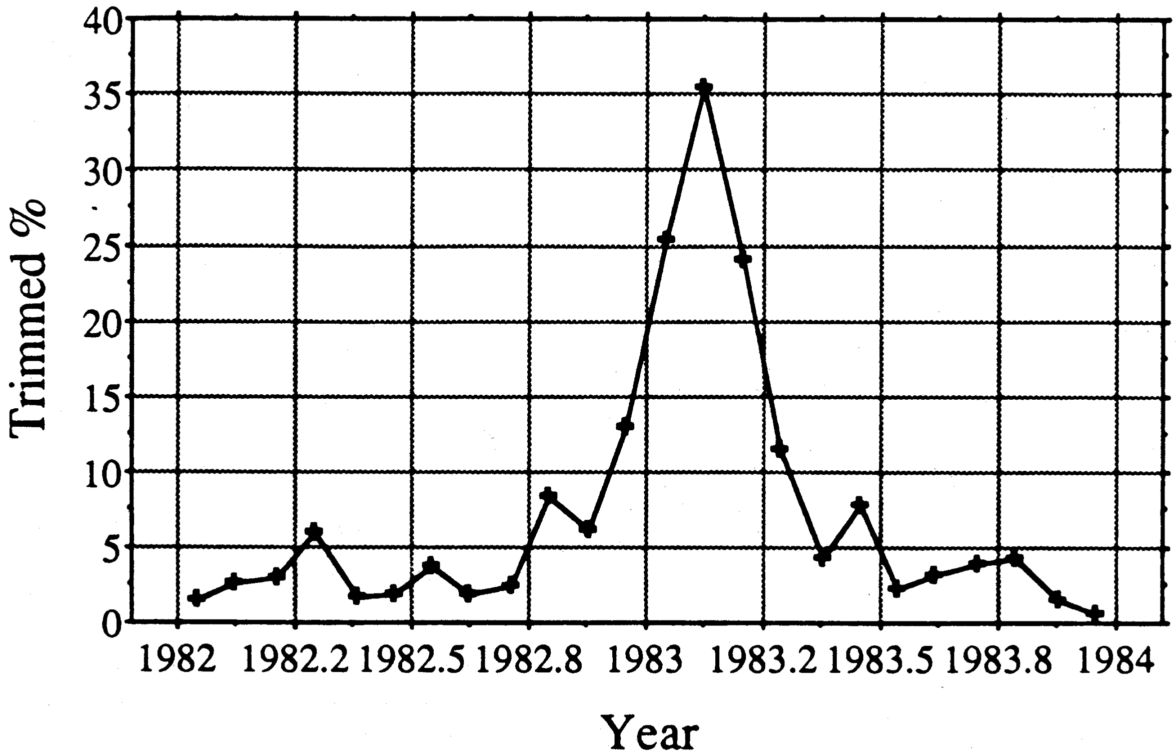

Figure 2. Losses due to trimming (in percent of total monthly SST observations) in ENSO 3 for January 1982 through December 1983.

Outliers can easily be detected in histograms of February 1983 data (Fig. 3). For instance, coding errors of +/-10°C (changing anomalies in Fig. 3a from +4°C to -6°C, or +1°C to +11°C, and codings of 0°C (anomalies of about -26°C in Fig. 3a) are obvious outliers, which, if not removed, would constitute Type II errors. They amount to about 2.7% of the untrimmed data, only a fraction compared to 32.7% of the same data that are trimmed erroneously (Type I errors) at the upper limit (compare Fig. 4a with 4b). Therefore, the untrimmed distribution between 0 and +6°C in Fig. 3a appears to be less skewed than the trimmed distribution between 0 and +4°C in Fig. 3b. The drawbacks of the present trimming limits are even more obvious in Fig. 4, where the trimmed distribution relative to the upper trimming limit (U) is truncated at OOC (Fig. 4b), while the untrimmed distribution appears “normal” out to about +3°C (Fig. 4a). Therefore, the present trimming process does not only remove obvious (coding) errors, as it is supposed to, but also a large number of realistic observations, which is the main concern here. Histograms like the ones in Figs. 3 and 4 capture this situation for homogenous regions such as ENSO 3, and help to differentiate between Type I and Type II errors. In other areas, this is not so easy.

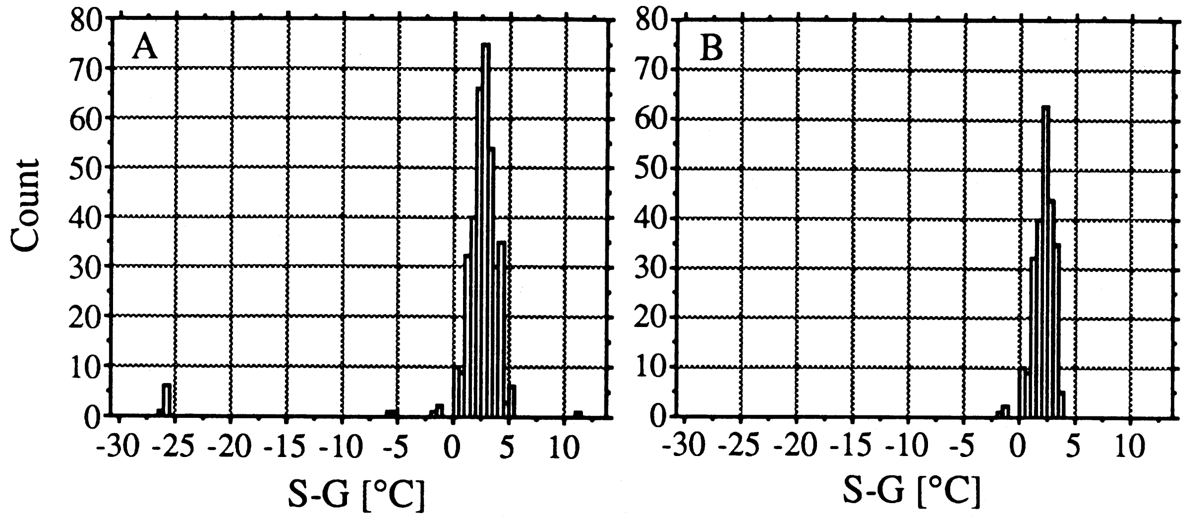

Figure 3. Histograms of ENSO 3 SST observations for February 1983. Frequencies are given for the anomalies of observed SST minus the long-term median value (G) for February 1950-79, computed for each 2°x2° square separately: (a) untrimmed, (b) trimmed. Bins are partitioned into 1/2°C intervals, with observations being greater than or equal to the lower boundary, and smaller than the upper boundary, starting at -30°C.

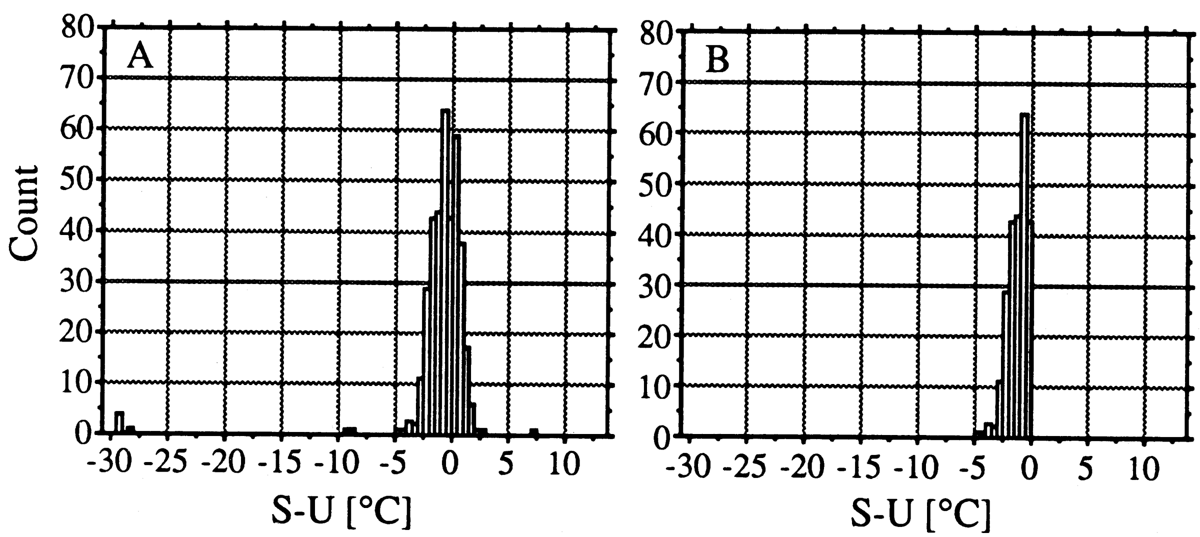

Figure 4. As in Fig. 3, except for subtracting the long-term upper trimming limits (U) rather than the long-term median (G) from each observation.

3. Alternative Trimming Solutions

Alternative trimming methods, namely inflated current limits, fixed 7°C limits, "Winsorization", floating two σ limits, and "human expert" screening are defined, discussed, and compared to no trimming and the current trimming approach for 1982/83 SST, based mainly on ENSO 3 data.

Alternative #0: Status Quo

Since extreme climate events are severely censored in this approach (ref. Fig. 4), it is argued that current trimming products should only be kept as "benchmark" values.

Alternative #1: No Trimming

Wolter et al. (1989) argue that old COADS data (prior to 1965) were sufficiently quality controlled before being keypunched so that one may be able to get by without trimming (aside from all the other error problems; ref. Trenberth et al., 1992). Unfortunately, this is out of the question with new data (especially post 79), mainly due to the large increase in error-prone telecommunicated data. If available, untrimmed summaries constitute a second "benchmark" set.

Alternative #2: Inflation of Existing Limits

This alternative (suggested by Roy Jenne, pers. comm. 1989) would have the advantage of easy implementation (trimming limits would only have to be multiplied by a simple factor [here: 5/3.5], not recalculated from scratch). However, it increases the risk of admitting too many "bogus" (+/-10°C) data (i.e., Type II errors) where σ is big, for instance in the Gulf Stream and Kuroshio regions. Nevertheless, for most regions of the World Ocean a slight inflation of the existing trimming limits would quite probably strike a better balance between Type I and Type II errors than the current procedure.

Alternative #3: Fixed 7°C Limits

Fixed centigrade limits about long-term means or medians are the simplest alternative. Bottomley et al. (1990) used +/-6°C as initial limits to screen SST anomalies about their long-term mean annual cycle with a space-time resolution of 1°x1° and pentads (five-day means). Problems with the determination of the exact size of the limits (here: 7°C) arise due to two competing factors: areas outside ENSO 3 may require limits larger than 7°C, but once you get close to 10°C, "bogus" data can sneak in easily (Fig. 3a). Again, areas with large spatial gradients and strong seasonal cycles are particularly vulnerable (ref. Trenberth et al., 1992). On the other hand, many areas never even come close to such anomalies, so 7°C could be too loose a restriction if the real anomalies never exceed, say, 3°C. Of course, allowing for geographical variation would forfeit the simplicity of fixed limits.

Alternative #4: Winsorization

This approach has been employed by Bottomley et al. (1990). After the initial screening described in alternative 3, it uses individual monthly quartiles rather than long-term trimming period limits, thereby embracing the idea of a variable climate. Within each box, it sets all anomalies below the first and above the third quartile (25%, 75%) equal to these quartile values. In other words, every observational value, whether an outlier or not (if within +/-6°C about the long-term pentadal mean), contributes to the monthly box mean, but may be strongly reduced in its anomaly. Although this is a robust method, computationally simple and applicable to all data, it suffers from the drawbacks of distrusting inherently 50% of all observations, and from largely removing the influence of the skewness on the monthly mean. In addition, it requires at least five individual observations per box to compute quartiles that clearly isolate outliers, a rare situation south of 20°N. In fact, erroneous sequences of data with low scatter are not eliminated unless outside the initial 6°C limits. This problem seems to arise mainly with poorly calibrated buoy data from areas with little independent data (pers. comm., Richard Reynolds, 1992).

Alternative #5: Floating 2 σ Limits - Individually Trimmed Months

Originally proposed by the author (Wolter, 1989), the "floating limit" approach trims the individual observations of each month and 2°x2° square about its median with limits of $+2*\sigma_u$/ $-2*\sigma_l$ (based on upper/lower σ_u/σ_l computed from the same month's sextiles). Adjusted limits on a month-by- month basis allow for extreme interannual variations to remain in the data and not be thrown out. In a situation with many observations per month and box, this alternative would be superior to the other ones, and leave roughly 95% of the data intact. However, it requires more than six observations to even catch a single outlier. As in Winsorization, it is vulnerable to sequences of nearly identical data that cannot be eliminated if erroneous, or, in turn, would yield limits that are too tight, and thereby trim too many reasonable observations. Such a particularly small range of sextiles occurs often enough in World Ocean SST (not in ENSO 3) to render this method less expedient than originally thought.

Alternative #6: Human Expert Screening

This approach was taken by the author for this particular case study only. Given the clear indication of outliers in Figures 3 and 4, such outliers were removed on an individual basis. In addition, continuity checks were employed in space and time, i.e., for jumps in SST anomalies relative to neighboring squares and compared to adjacent months. Discontinuities of more than two standard deviations were flagged as outlying observations and removed. In this particular case, relatively few such cases were found. Alternative 6 is hampered by the limited availability of data for comparisons in most regions south of 20°N, including ENSO 3.

Figure 5 (below) shows the SST anomalies based on alternatives 0 - 6 for ENSO 3 in 1982/83. Note the peak in December 1982 for ΔSST, while the biggest discrepancies between the current trimming results and the new trimming results occur in February 1983 (along with the highest number of conventionally trimmed observations; ref. Fig. 2). Given the small scatter of alternatives 2 - 6 through the full two years, it is tentatively concluded that the true temperature anomaly time series would be close to their average. In ENSO 3, any of the presented alternatives 2 - 6 would improve upon the currently available trimmed COADS product.

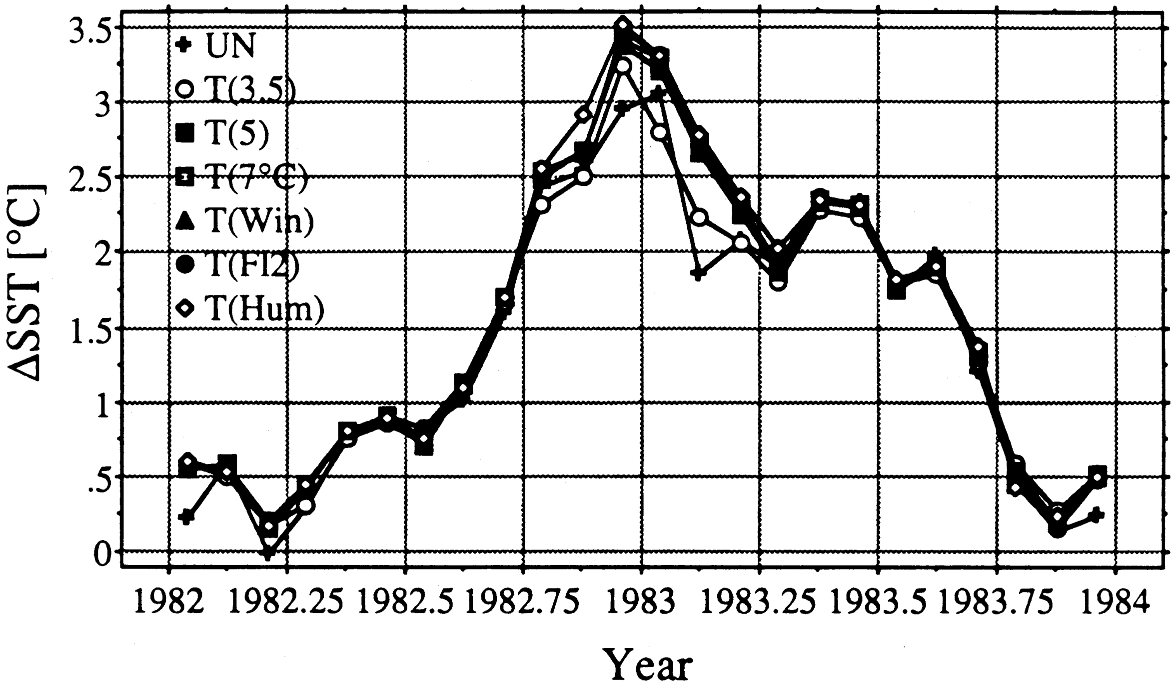

Figure 5. Monthly trimmed averages of SST anomalies for Alternatives 0 - 6 in ENSO 3 for January 1982 through December 1983. Anomalies are computed with respect to long-term (1950-79) median values. The alternatives are: (0) conventional trimming [T(3.5)]; (1) no trimming UN]; (2) inflated trimming limits from 3.5σ to 5σ [T(5)]; (3) trimming with respect to fixed 7°C limits about the long-term median [T(7°C)]; (4) winsorized data after eliminating observations with more than 7°C anomalies from the long-term medians [T(Win)]; (5) trimmed data with respect to 2σ limits based on individual monthly medians and sextiles [T(FL2)]; (6) trimmed data according to human expert judgement [T(Hum)].

The geographic distribution of SST anomalies in February 1983 is displayed below in Figure 6. A number of anomalies over +3°C are completely lost in the current trimming procedure (Fig. 6a), while alternative 2 (Fig. 6b, or any of the other alternatives 3 - 6) retains them in a realistic manner. ENSO 3 SST anomalies in February 1983 average about 0.5°C higher with alternatives 2 - 6 than with alternative 0 (Fig. 5), while the highest local anomalies are often more than 1°C higher in the revised trimming scenarios than in the original trimmed data (Fig. 6c). These local extremes, along with total losses in alternative 0, mainly occur in the western and central parts of ENSO 3 where low observational densities are common (often just one or two observations per month and 2°x2° square).

In sum, wherever interannual variability is large compared to within-monthly variability or observational noise, the present trimming procedure underestimates the true size of extreme interannual events such as the 1982/83 ENSO event. Central Pacific SST is particularly strongly affected by this feature (1982/83 ΔSST reduced by up to and over 1°C. Alternative trimming methods, such as increased existing limits (Alternative #2), fixed 7°C limits (3), Winsorization

(4), and floating 2σ limits (5) all improve the trimmed product considerably and are consistent with each other for the particular event (1982/83) and region (ENSO 3) discussed here.

However, alternative 2 (3), although quite appropriate in the ENSO 3 setting, appears to be less optimal for situations with larger (smaller) "natural" variability. Alternatives 4 and 5, although superior in areas with sufficient observational density, become less practical south of 20°N where fewer observations are taken. Alternative 5 has the particular tendency to remove too many observations in the presence of low variability within a month. In addition, all discussed trimming alternatives are ill equipped to handle runs of poorly calibrated (buoy) data. Alternative 6 is discarded as impractical for areas with low observational densities. The present trimming approach (0), although discredited for tropical SST, is less problematic for other variables, while no trimming (1) is a viable alternative for old data (<1965).

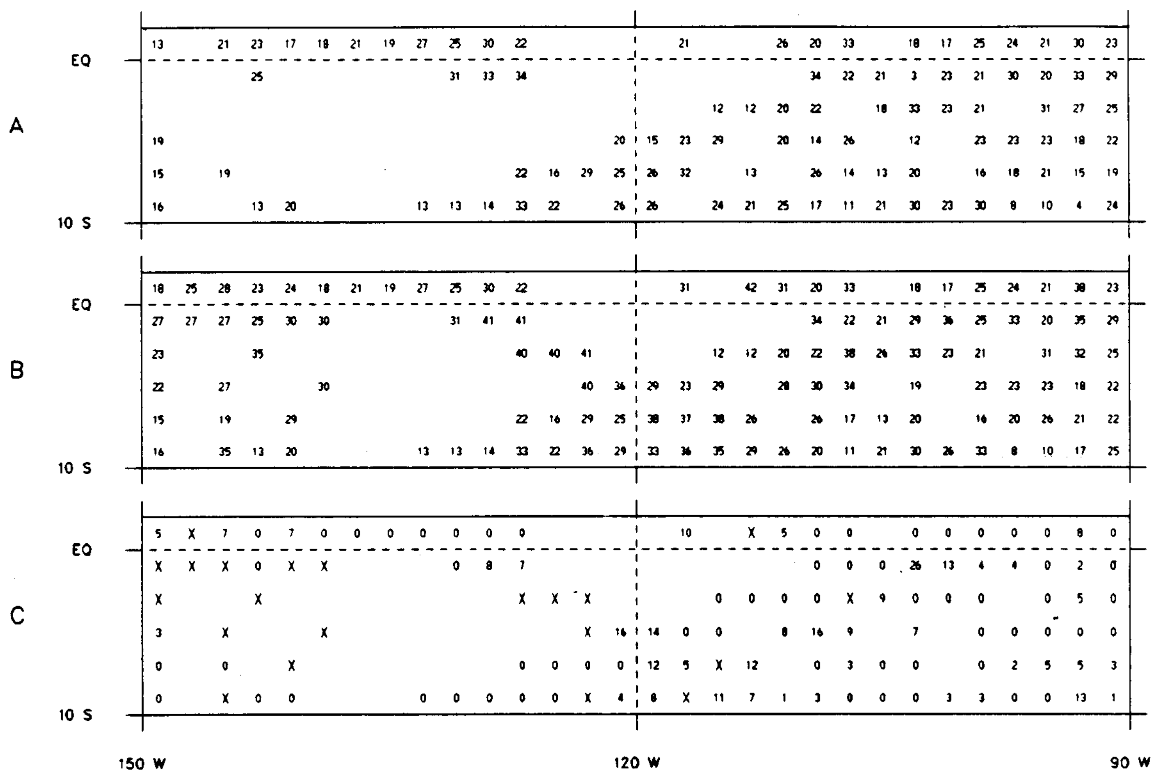

Figure 6. Spatial distribution of SST anomalies [in 1/10°C] in February 1983 within ENSO 3 for: (a) T(3.5) trimming, (b) T(5) trimming, and (c) the difference [T(5) - T(3.5)]. If data made it through the wider trimming limits in (b), but not in (a), it was marked X in (c).

4. Outlook and a new general alternative trimming method

Trimming should not be considered in isolation. It is an important part of comprehensive quality control procedures applied to COADS. Future improvements in space and time resolution and platform separation will have to be balanced against the penalties of reduced sample sizes. Observed geophysical data are characterized by our inability to control or repeat the events we are monitoring. This implies that there is no way of separating true statistical outliers from true climate extremes with absolute confidence. Multimodal observational distributions or those with prominent tails (i.e., high kurtosis) are the most difficult to trim. Resorting to statistical methods to minimize Type I as well as Type II errors, we can never be completely sure that we got rid of them.

After this paper was given, one last alternative trimming approach was conceived of. It is described below, since it appears to be the most balanced general solution to the encountered problems:

Alternative #7: Trimming separately within-monthly scatter and inter-annual variability

The interannual variance of a given box can be separated from the within-monthly variance. In other words, the observed median of a given month and its anomaly from the long-term climatic median is considered separately from the within-monthly scatter around that median, so that certain ranges of either are only allowed. This approach combines ideas from alternatives 4 and 5, i.e., the Winsorization and floating 2σ limits alternatives.

It is practical to start this procedure with the trimming of the within-monthly scatter. In order to avoid Type II errors, the allowable range within each month and box should be on the order of +/-$2*\sigma_w$ (σ_w stands for the within-monthly standard deviation). Analogous to the present trimming procedure, σ_w is computed on a decadal basis, then averaged for three decades and smoothed across a “hypercube” of three adjacent months and up to nine adjacent squares (Slutz et al., 1985). Compared to alternative #5, a long-term measure of the within-monthly scatter is used rather than being at the mercy of the scatter of the individual month at hand, which can be quite misleading, especially if only a few observations constitute this scatter (ref. discussion of 5). In fact, even just two observations in a given box and month can thus be trimmed. If only one observation is left, it is trimmed as an observational median (see below). By using a permissible range of +/-$2*\sigma_w$, about 95% of normally distributed data will be preserved, less than presently allowed, but considerably more than in alternative 4 (or, in practical terms, alternative 5). Since about 1979, clear outliers, such as encountered in GTS data with altered leading digits, occur on average between 1% and 3% of the time, which is why one cannot relax these trimming conditions much further (the exact numerical range will be fine tuned in a series of experiments with COADS data). Furthermore, trimming at the tails of each month’s distribution (i.e., removing successively the most extreme observations until the +/-$2*\sigma_w$ range condition is fulfilled) only changes the individual monthly median significantly if at least half of the observations are out of bounds.

After the individual observations have been trimmed, the interannual scatter of the observational medians (based on the remaining data) is trimmed in a similar fashion. In order to avoid the Type I errors illuminated in section 2, the admissible range of the observed medians should be fairly large, say, $+4*\sigma_{iu}/-4*\sigma_{il}$ (σ_i stands for the interannual standard deviation, with the possibility of asymmetric σ_{iu}/σ_{iu} to account for a skewness in the distribution of the medians). Again, σ_i is computed on a decadal basis, then averaged for three decades and smoothed across three adjacent months and up to nine adjacent squares. If a given median falls outside this wide range, it could be either removed wholesale or flagged as suspicious and await further tests (such as position checks). As in the first part of this combined trimming procedure, the exact optimal range for the observational medians will be determined in a suite of tests with COADS data in different regions of the World Ocean. For instance, in the tropical Pacific studied here, σ_W is relatively small, but σ_i can be quite large (ENSO), so one would end up with comparatively loose interannual, but tight within-monthly limits (see Fig. 7 below). On the other hand, the Gulf stream and Kuroshio regions have large synoptic variability plus strong spatial gradients, but comparatively less interannual variability, so one could use tighter interannual, but looser within-monthly limits.

To illustrate the difference between within-monthly (σ_W) and interannual variability (σ_i), Fig. 7 presents a first look at the ratio between these two quantities in the domain of Fig. 1.

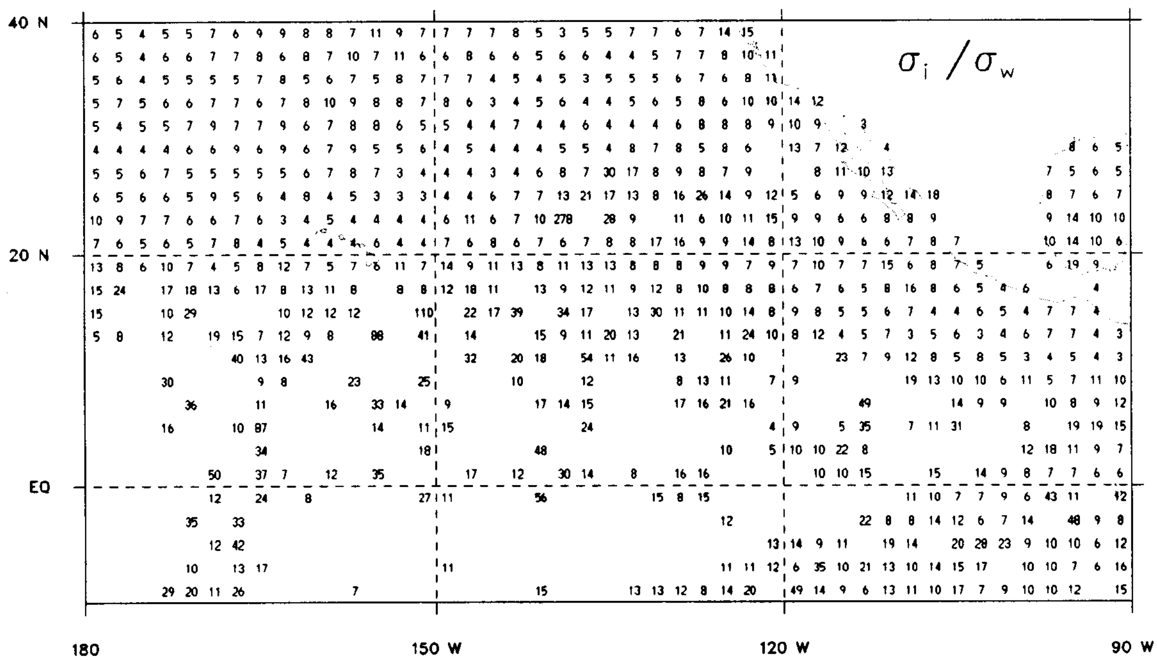

Figure 7. Approximate ratio [in 1/10] of interannual standard deviations (σ_i) to within-monthly standard deviations (σ_W) for February 1950-79. The domain is identical to the one in Fig. L While σ_i's are computed from sextiles of the <u>trimmed</u> 2°x2° February averages within this time period, σ_W's are the residuals based on the squared difference between the long-term total standard deviation (based on sextiles) of all <u>untrimmed</u> observations within February 1950-79 and the aforedescribed σ_i's.

Since σ_i was computed from the <u>unsmoothed</u> trimmed summaries, the resulting map (Fig. 7) is noisier than a true ratio between smoothed σ_i over σ_W would be. They typical size of this ratio is less than 1.0, even less than 0.5, north of 20°N, while 2.0 is commonly approached or even exceeded near the Equator. In other words, σ_i is often four times as important in the ENSO regions than in the subtropics compared to σ_W. Tests for implemention of this new trimming approach under way.

Acknowledgments: This research has been funded through the NOAA Climate and Global Change program. Mei Yuen Liu has helped with the data processing and Mike Timlin with the figures. discussions with Scott Woodruff and Henry Diaz are gratefully acknowledged.

References

Bottomley, M., Folland, C. K., Hsiung, J., Newell, R. E., and D. E. Parker, 1990: *Global Ocean Surface Temperature Atlas.* MIT Press, 20 pp. plus 313 maps.

Slutz, R. J., Lubker, S. J., Hiscox, J. D., Woodruff, S. D., Jenne, R. L., Joseph, D. H., Steurer, P. M., and J. D. Elms, 1985: *Comprehensive Ocean-Atmosphere Data Set. Release 1*. NOAA Environmental Research Laboratories, Boulder, Colorado, *268 pp.*

Trenberth, K. E., Christy, J. R., and J. W. Hurrell, *1992:* Monitoring global monthly mean surface temperature. *Journal of Climate,* (in press).

Wolter, K., *1989:* Trimmed COADS - problems and solutions. *Proceedings of the 14th Climate Diagnostics Workshop, La Jolla, California, October 1989,* pp. 228 - 233.

Wolter, K., Lubker, S. J., and S. D. Woodruff, *1989:* Trimming - a potential error source in COADS. *Tropical Ocean- Atmosphere Newsletter,* **51** (Summer 1989), pp. 4 - 7.

Woodruff, S. D., Slutz, R. J., Jenne, R. L., and P. M. Steurer, *1987: A* comprehensive ocean-atmosphere data set. *Bulletin of the American Meteorological Society,* **68**, 1239 - 1250.

Woodruff, S. D., Lubker, S. J., and M. Y. Liu, *1992:* Updating COADS - Problems and Opportunities. *Proceedings of the International COADS Workshop, Boulder, Colorado, 13 - 15 January 1992.* H. F. Diaz, K. Wolter, and S. D. Woodruff, Eds., NOAA Environmental Research Laboratories, Boulder, CO (this volume).