Long-Term Trends in Surface Temperature over the Oceans

T. P. BARNETT

Climate Research Group, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093

(Manuscript received 16 June 1983, in final form 23 September 1983)

ABSTRACT

A limited comparison over the Northern Hemisphere oceans has been made between sea surface temperatures obtained from "Marine Decks," air temperatures over the ocean obtained from the same decks, and the historical file of hydrographic data. The intercomparison of these data suggest the following conclusions.

1) The SST observations have been contaminated by a systematic conversion from bucket to injection measurements. The bias so introduced may constitute as much as 30 to 50% of the observed change in sea surface temperature since the turn of the century.

2) The same bias effects are apparent in data sets that are alleged to contain bucket measurements of sea surface temperature only.

3) The behavior of the temperature field over the ocean appears to have significant and substantial differences from the behavior of estimated temperature changes over the Northern Hemisphere land masses. It seems clear that a reliable estimate of hemispheric or global temperature cannot be made without including adequate coverage of the ocean regions.

4) Many of the data suggest a shift in mean state of the sea surface temperature field of certain regions. It appears that this change is partially real and partially due to the merging of rather different types of data.

5) All estimates of pentad average temperature since 1900 are attended by relatively large standard deviations. This fact makes definitive discussion of pentad-to-pentad or decade-to-decade changes in hemispheric and global temperature difficult, if not impossible. Different methods of calculating the uncertainty in pentad averages could effect this conclusion.

1. Introduction

Long term trends in hemispheric and/or global air temperature are a subject of increasing interest and importance. Such changes are closely inspected for effects of aerosols, CO_2, volcanoes and so on. However, some argue that adequate estimates of "global temperature" change, at least, have yet to be made (e.g., van Loon and Williams, 1976; Barnett, 1978). A major problem they raise is adequate sampling of air temperature change over the 72% of the earth's surface covered by ocean. It can be speculated that the problem, while still present, may not be so severe in the Northern Hemisphere due to the small relative ocean area.

Beyond the seasonal averaged time scale, air temperature changes over the oceans are closely related to sea surface temperature (SST) changes (Cayan, 1980). So the problem of estimating air temperature over the ocean is usually posed in terms of determining the change in SST. A recent attempt in this direction has been made by Paltridge and Woodruff (1981) who used ship reports of SST largely from the Historical Sea Surface Temperature Data set (HSSTD) to estimate temperature changes over the oceans. Their work had several potential analysis problems but nevertheless represented a first attempt to develop information on long term SST changes over large areas of the global oceans. Their results when compared with previous estimates of global temperature based on land data, suggest the long-term temperature trends over the ocean differ significantly from those over land. The clear implication is that long-term temperature changes based on "land only" data are not necessarily indicative of a global change.

The main source of SST observations used by Paltridge and Woodruff, and others, to investigate SST changes to the turn of the century and beyond are the historical archives of surface ship observations. The millions of observations from the world's oceans contained in this archive suggest it is virtually the only hope of getting a globally comprehensive view of long term surface temperature changes. Indeed this data set has already told us nearly everything we now know about short-term changes in ocean climate. However, is the data in this archive good enough to investigate small, long-term trends in temperature over the ocean? That question is investigated in this paper.

The main problems addressed here are: 1) estimating the importance to long-term temperature trends of using SST data obtained by different measurement techniques when the blend of measurements has changed over the last 80 years, 2) determining if measurements of air temperature from ships really are identical to SST and, if so, do the two measurements of surface temperature agree with long-term temperature measurements over land and 3) is a shift of climate regime really evident in the ocean data.

2. Data

Archived ship observations from the three Northern Hemisphere oceans will be used in this study. This ensemble of data will be referred to here as the "Marine Deck" for simplicity. However, the origin of the sets is important and is discussed briefly below.

Data from the Indian Ocean was obtained directly from the Koninklijk Nederlands Meteorologisch Instituut for the period 1861–1960. This set of observational data (not the "summary set") represents the Dutch contribution to the Historical Sea Surface Temperature Data (HSSTD) project. The data are thought to be screened to eliminate most SST data not taken by the bucket method. This is an important point for later analysis. After 1960 the Indian Ocean data come from the National Climatic Center's "Surface Marine Observations" (NCDC, 1968) and presumably contain all observations regardless of how the SST data were taken.

Data for the Pacific Ocean were available to us exclusively from the Surface Marine Observations and include bucket and non-bucket reports. The description of the data and the individual decks that make up this set (e.g., NCDC, 1968; Verploegh, 1967) clearly show that the majority of the data prior to the early 1940s comes from European sources. After the early 1940s the Surface Marine Observation set relies heavily on U.S. Merchant Marine observations. The former data are thought to observe a significant part of their SST data by bucket methods while the latter are thought to come largely from injection temperatures (see Section 3).

Data for the Atlantic Ocean were obtained from the National Climatic Data Center and were copies of the German HSSTD observational set (TD9774, R. Quayle, personal communication, 1983). These data also have supposedly been filtered to eliminate most SST reports from non-bucket sources.

3. SST problems: Effects of instrumental bias

a. The problem

Observations of SST from ships of opportunity today, and over at least the last three decades, are thought to come largely from thermometers located in the ships' water cooling intake line, and hence are called injection temperatures. It is well known that the temperatures so obtained are biased high (e.g., see Roll, 1965 for a complete discussion).

The above fact is shown graphically in Fig. 1 where two long-term climatologies for the North Pacific SST field are compared. The climatologies were constructed in an identical manner except one field used standard

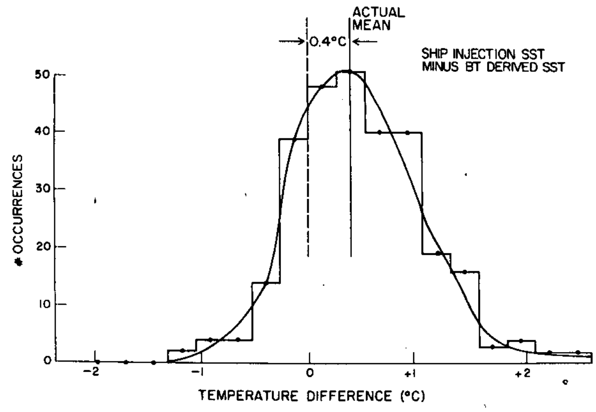

FIG. 1. The distribution function for the difference between two long-term climatologies for the North Pacific SST field.

ship reports (largely injection temperatures) while the other was constructed solely from surface bucket reports from approximately 10^5 bathythermograph (BT) observations in the same area/time span. The two climatologies were subtracted and the combined distribution function of the difference field for all grid points/months plotted (see Barnett and Ott, 1976, for more details). The bias seen in Fig. 1 is 0.4°C, with ship reports higher than BTs. Others have estimated this bias in very different ways and come up with essentially the same number; e.g., Saur (1963) found 0.7°C. Tabata (1978) obtained 0.3°C and James and Fox (1972) found 0.3°C. There seems little doubt about the reality of the bucket–injection bias.

In olden times SSTs were supposedly reported from bucket temperatures. Bucket temperatures have potential problems of their own that will tend to make them biased low (Roll, 1965; Verploegh, personal communication, 1983) thereby compounding the problems noted in this paper.

Since sailing ships largely left the fleet, circa 1900, (Moyse–Bartlett, 1946) one wonders how buckets/injections mixed over the years as more opportunity arose to replace buckets with a much more easily obtainable injection value—imagine the temptation at night and/or in rough seas. There seem to be no good records to decide this crucial question. It is known that Surface Marine Observations are a blend of SST measured by the two different techniques. It is often stated that the HSSTD sets have few injection temperatures; but is this a provable fact or simply an educated guess or a statement of faith? The physical evidence, or people, needed to settle the matter, seem to have disappeared into history (Quayle, personal communication, 1983). Finally, beginning sometime around World War II, the mix of measurement techniques seems to have become a more accepted practice so that data since, say, the 1950s apparently contains both types of measurements.

A scenario, based on the facts and unknowns cited

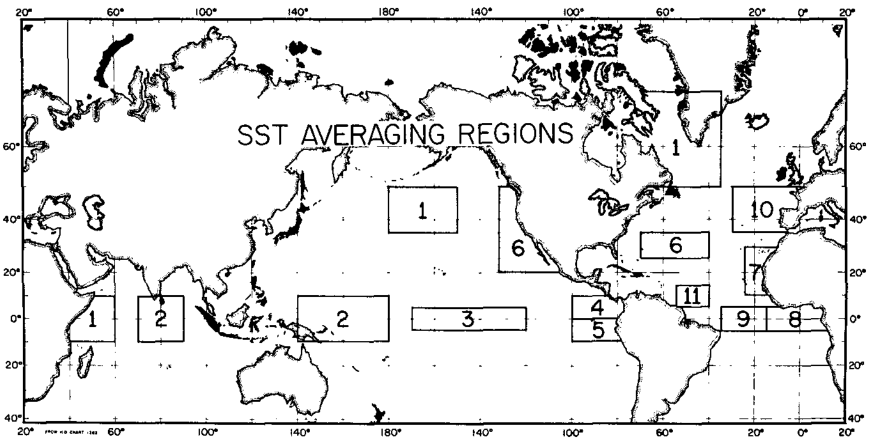

FIG. 2. Large area averaging regions used in this study.

above, can easily be constructed to explain the apparent increase of SST seen since the turn of the century (Section 3b). Suppose we begin in 1900 with only bucket reports as the sole source of SST data, then gradually increase the number of injection observations relative to buckets until the latter has largely replaced the former data type. Assuming an approximately linear increase in mixing of observation types in time will give an apparent increase of SST since 1900 of roughly 0.3–0.7°C. This is essentially the value rate found by Paltridge and Woodruff (1981) and stated in the next section of this paper. Yet the whole apparent change could be completely false. Further problems can arise if the two data types were not gradually mixed [cf. Section 5, item (c)].

b. Estimates of SST change

The idea of this section is to obtain area averaged, long time series of SST from a variety of regions[1] of the equatorial and Northern Hemisphere ocean by two independent means and then to compare them. The first set of series was constructed from the Marine Deck's ship observations and will be denoted by SST_S. These series presumably could follow the scenario noted above. A second set of time series was constructed for the same regions using all available hydrographic data in those areas. The "hydro" data (SST_H) is high quality (±0.02°C) and not subject to changing procedures, instruments, etc., that would induce a trend. Finally, time series of ship-reported air temperature for the same regions were developed from the Marine Deck and compared with SST_S. Details of these analyses are as follows:

SST from Marine Deck. All raw ship reports were extracted from the data sources of Section 2 for the large averaging areas shown in Fig. 2. Temperature variations in these regions were found to be highly coherent based on an EOF analysis, thus justifying the averaging procedure. The data were arranged by 1° latitude × longitude squares, month and year. Long term means were estimated for each month/square and anomalies were formed for each square/month/year. These values were used in another pass through the raw data to eliminate bizarre values, and the means and anomalies were recomputed. This 1° grid data where and when it was available was then averaged over the area and over the year to give a yearly anomaly for each large area. During any one month/year a given large averaging area could be represented by a relatively few 1° grids and this, no doubt, adds considerable noise to the large area averages. The potential effects of different observation times through the years is also a hazard that is partially explored in the Appendix.

SST from hydrographic observations. All available hydrographic cast data for the regions shown in Fig. 2 were extracted from the computerized historical files of NODC kept at the Scripps Institution of Oceanography. The data were arranged by the same 1° latitude × longitude squares, month and year as were used for the SST_S data. The long-term means for each month/square from the SST_S analysis were subtracted to form anomaly series. The resulting anomalies (SST_H) were averaged into yearly values for each large area. Note the estimated annual cycle from the SST_S data may be biased if the scenario of Section 3a is correct. This bias will show as a constant offset in the SST_H time series.

Air temperature from Marine Deck. Development of the air temperature (T_A) anomaly series followed the same procedures as used for the SST_S series. The measurement of air temperature aboard ship is at-

[1] A major, potential shortcoming of this study is the lack of full ocean coverage. However, the uniformity of the results suggest the "representative" area approach gives answers that will not differ radically from those obtained by a "full ocean" study.

tended by as many, if not more, problems than the SST. Changes over the years in thermometer enclosure style and location, diurnal effects, etc., are potentially quite severe (cf. Goerss and Duchan, 1980; Husby, 1980; Reynolds, 1982; Seguin and Garstang, 1971, among others). The effects of changing from thermometers in enclosures to sling thermometers may be very significant. In this paper we tacitly assume the air temperature data are "correct." It may come as some surprise that these data will yield results in relatively good accord with other (independent) data sets. However, the discussion of the Appendix shows the T_A to be attended by potentially serious temporal sampling problems so the "correctness" assumption must be considered during the following discussion.

c. Results

Examples of a regional comparison between SST_S and SST_H for a data rich area in the eastern Atlantic are shown in Fig. 3a. Similarly, a comparison between SST_S and air temperature (T_A) for a data rich area of the North Pacific is shown in Fig. 3b. These results, and others like them, generally show that the regional averages were too noisy, by themselves, to allow a good comparison of the variables. This was especially true for the SST_H.

The noisy nature of the data was partially overcome by averaging over time and entire ocean basins and then all basins for Figs. 4–7. The averages representative of individual oceans (Figs. 4 and 6) were obtained as a simple linear sum of the regions in a given ocean with each region weighted by its area. The values representative of the Northern Hemisphere ocean were the average of the three individual ocean series weighted by the relative areas each ocean occupied in the Northern Hemisphere (Figs. 5 and 8). This procedure will clearly favor effects in the North Pacific Ocean.[2] The yearly values of oceanic SSTs and T_A were then averaged by pentad, noting with parentheses pentads composed of two years or less of data. Similarly, pentad values for the Northern Hemisphere ocean consisting of less than 2 years and/or contributions from one ocean only are noted in parentheses. Also shown are average values of pentad standard deviations obtained by 1) estimating the standard deviation of the annual averages about each pentad mean and, 2) averaging these individual standard deviations over all pentads. This averaging procedure gives a measure of "typical" variation; not the variation associated with each pentad average.

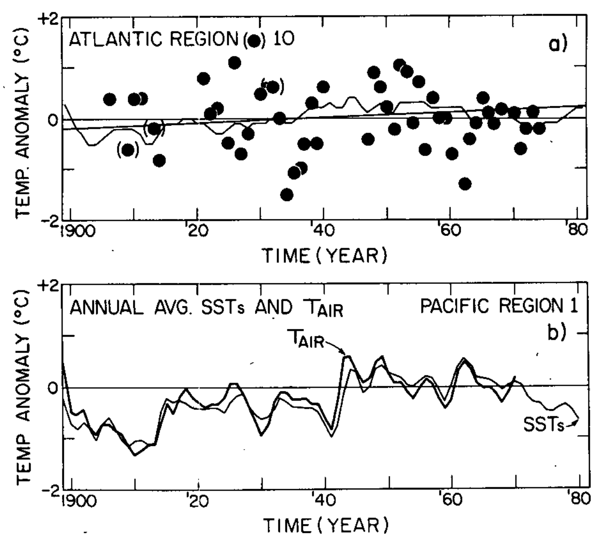

FIG. 3. (a) A typical regional comparison between sea surface temperature from the Marine Deck (solid line) and similar estimates from the hydrographic data (dots) for the data-rich region 10 in the eastern Atlantic. The best fit trend line to the Marine Deck data is also shown. Hydrographic annual averages composed of 2 or less observations are shown in parentheses. (b) A similar comparison between SSTs from the Marine Deck and air temperature from the Marine Deck for a data-rich area in the central North Pacific (region 1). All anomaly values are shown in °C.

The above illustrations lead to the following conclusions regarding the agreement between SST_S, SST_H and T_A.

1) The general positive trend of SST_H in the Pacific and Atlantic Oceans is somewhat similar to the SST_S (Fig. 4). However, the Indian Ocean data sets do not appear similar. The relatively large error bars on the Marine Deck data and hydrographic data do not permit a quantitative confirmation of the apparent agreement/disagreement, except in the Indian Ocean where the divergence of the two data sources does appear significant in recent times.

2) The grand average of SST_S for the Northern Hemisphere oceans (Fig. 5) agrees rather well with similar estimates by Paltridge and Woodruff (1981) although the magnitude of the signal is less by roughly 0.2 to 0.3°C. However, ignoring the 1912 pentad for the moment since those data come exclusively from the North Atlantic, the hydrographic data only weakly show the same increase. The best that can be said is that the hydrographic data "suggestively" supports the SST_S due exclusively to the 1962 and 1967 pentads data from the North Pacific. The correlation between SST_S and SST_H for the period 1922–77 is 0.42, a value not significant at the 90% level, even assuming independence of successive pentads (a poor assumption). The size of the error bars, plus the variability of the time series themselves, again make a more quantitative statement impossible.

3) The last 5 pentads of record (1957–77), a time when SST_S probably came largely from injection reading and when there was a high density of hydro data,

[2] It is important to remember that these averages are not made over the full area of the ocean(s); rather they are proxies of the full ocean averages.

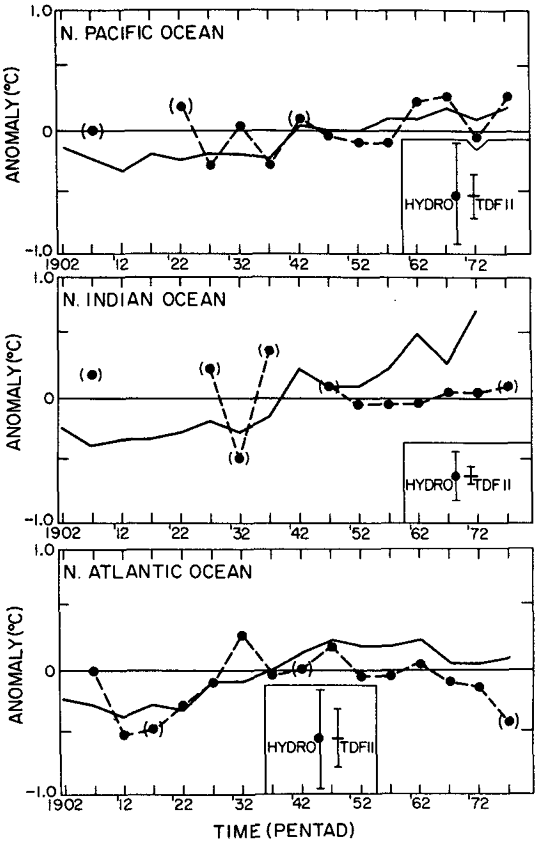

FIG. 4. Time series of sea surface temperature anomalies from the Marine Deck (solid lines) and the hydrographic files (dashed lines) representative of the three Northern Hemisphere oceans (°C). The error bars refer to ±1 "typical" standard deviation computed from the five yearly values that were used to obtain the pentad average. Dots in parentheses denote two years of data or less.

were investigated quantitatively for systematic differences that appear in Figs. 4 and 5. The average difference of $SST_S - SST_H$ for the Atlantic, Pacific, Indian and Northern Hemisphere Oceans were +0.3°C, +0.0°C, +0.5°C and +0.15°C, respectively, with SST_S being warmer. The Pacific result, because of its large area, brings down the Northern Hemisphere ocean estimate. At any rate, the difference is uncomfortably close to the magnitude excepted from instrumental bias alone. The same procedure was repeated for Pacific region 2, Atlantic regions 1 and 10, for the 1930s, a time of relatively good hydrographic coverage in these areas. The differences were −0.04, −0.40, −0.15, respectively, with SST_S being lower in all cases. The lower values, particularly at high latitudes, may represent the negative bias problem associated with bucket data (see Roll, 1965; Verploegh, personal communication, 1983). At any rate, the absolute difference between the 1930s data and the 1960–70s data suggests SST_S has increased relative to SST_H by roughly 0.2–0.4°C. The limited, noisy nature of the SST_H data preclude a more exacting comparison.

4) The same approach was taken to evaluate the differences between SST_S and the equally numerous T_A (Figs. 6 and 8). The results suggest that during the period 1900–40 the SST_S were less "colder" than T_A by 0.04, 0.11, 0.10, 0.10°C for the Indian, Pacific, Atlantic and Northern Hemisphere oceans, respectively. For the period 1945–75, the SST_S were greater "warmer" than T_A by 0.04, 0.11, 0.08, 0.09°C for the corresponding ocean areas. Thus the SST_S appears to have increased relative to T_A over the oceans by approximately 0.2°C between the early part of the century and recent times; a result in independent agreement with 3) above. Höflich (1973) found a similar result (0.4°C) for the tropical Atlantic. These values are close to those expected of a large (percentage-wise) switch from buckets to injection temperatures.

5) The difference between pentad values of SST_S and T_A anomalies for the Northern Hemisphere Oceans is shown in Figure 7. Also shown is an apparently *a priori* correction scheme to SST_S values suggested by Folland and Kates (1982). The schematic nature of the change in the difference is certainly supportive of the scenario offered in Section 3a. The rather abrupt, discontinuous nature of the change is dictated largely by events in the Pacific (because of its relatively large area) and will be discussed in the next section.

As an alternative explanation, let us suppose for the moment that both SST_S and T_A data are correct. If the long-term means are equal then Fig. 7 implies a sensible heating of the ocean by the atmosphere between circa 1900–40. Since 1940 the reverse situation would hold. These relative rates of heating/cooling, if real, would lead to SST changes larger than observed. The Fig. 7 difference also suggests a change in latent heat flux of order 20–40% between the two time periods (taking typical mean values of SST, T_A and wind and assuming other pertinent atmospheric parameters to be constant). Such changes in latent heat flux, if real, should have had a noticeable impact on both the atmosphere and global water budget. Finally, the veracity

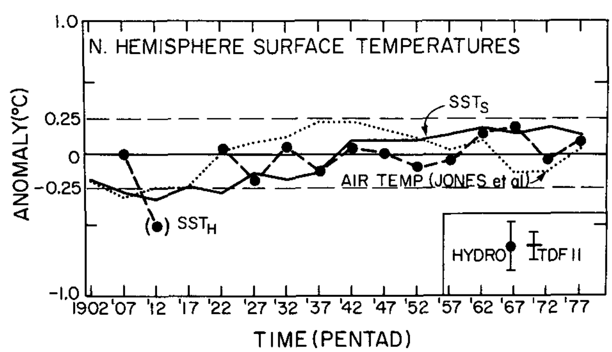

FIG. 5. Time series by pentad average of anomalies of SST from the Marine Decks versus SST from the hydrographic data file. The series are representative of the Northern Hemisphere oceans. Also shown for comparison is an estimate of Northern Hemisphere air temperature, from Jones *et al.*, (1982) based mainly on land data. See legend in Fig. 4.

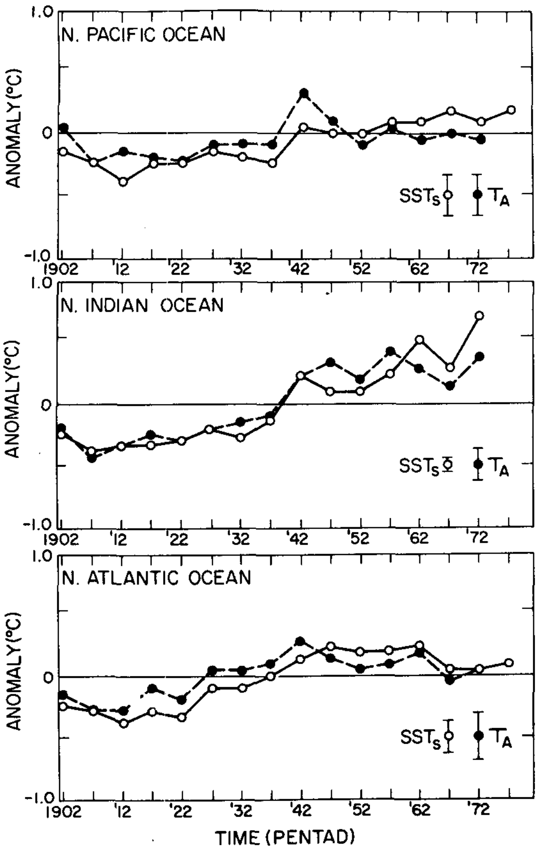

FIG. 6. Time series of sea surface and surface air temperature anomalies over the oceans from the Marine Decks. The data are representative of the three Northern Hemisphere oceans as indicated. See legend in Fig. 4.

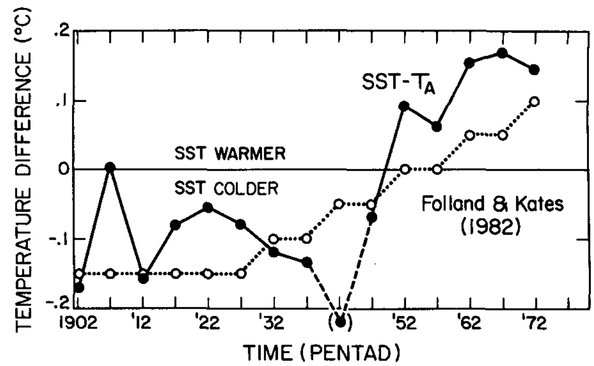

FIG. 7. The difference between SST from the Marine Deck and air temperature from the same data source by pentad. The *a priori* instrument bias corrections suggested by Folland and Kates (1982) are shown by circles and dashed lines.

of both T_A and SST_S is difficult to rationalize in consideration of Fig. 1 and other studies like it.

6) Note that the differences between SST_H/T_A and SST_S appear in all oceans. Thus the supposedly "buckets only" HSSTD sets (Indian/Atlantic) apparently have the same data discrepancies as do the Surface Marine Observations (Pacific) which admittedly includes both types of SST measurements. Alternatively, both data sets could be correct but then how does one rationalize results obtained by many workers regarding bucket-injection bias (cf. Fig. 1)? It appears more justification will be required before one can accept the claim that the SST data in the HSSTD set are not liable to instrumental bias effects.

In summary, two independent data sets suggest that estimates of SST_S since 1900 are subject to differential errors due to instrumental bias. The sense of the errors and the way they may have entered the data set would lead one to assume an increase in SST. The magnitude of these errors is roughly 30–50% of the computed increase since the early 1900's. However, since two independent data sets show some increase also, it appears that part of the SST_S increase is also real. Unfortunately, the magnitude is relatively uncertain.

4. Implications for "Northern Hemisphere"/"Global" temperature estimation

The Northern Hemisphere ocean series for SST_H, SST_S and T_A are compared in Figs. 5 and 8 with an estimate of Northern Hemisphere air temperature from Jones *et al.* (1982). There is a substantial quantitative difference between the data obtained over the oceans and that obtained over land, particularly between 1920–40. This difference was noted by Paltridge and Woodruff (1981). They thought the problem might be due, in part, to the SST_S bias effects.

It is now rather clear that the land–ocean temperature difference is not due entirely to instrumental effects since it is seen in two independently measured ocean data sets. While the magnitude of ocean changes is uncertain, the phasing of both ocean series generally agree with each other at least up to the 1950s when

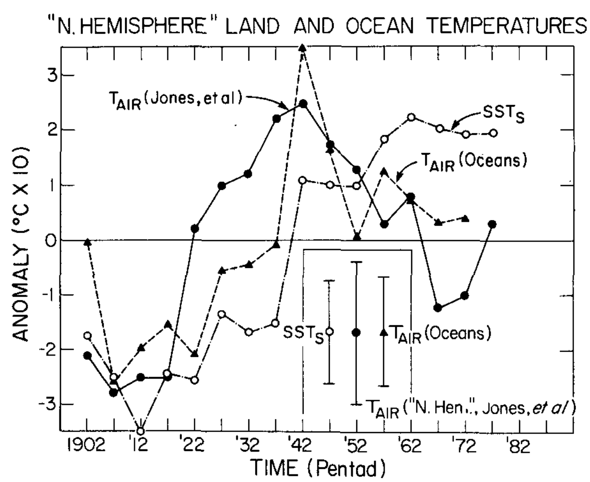

FIG. 8. Time series by pentad average of anomalies of SST from the Marine Decks versus surface air temperature from the same source. The series are representative of the Northern Hemisphere oceans. Also shown for comparison is an estimate of Northern Hemisphere air temperature, from Jones *et al.* (1982) based mainly on land data. See legend in Fig. 4.

the apparent bias in SST_S seems to preclude its usefulness without some type of correction.

Based on the data from 1900 to 1950, it appears impossible to discuss either Northern Hemisphere or global temperature changes without including changes over the ocean regions. It, therefore, appears that much of the work done in this area to date that claims to represent global changes might more accurately be described as representing changes over the land masses.

Developing a true global average temperature will be difficult using conventional measures if the results of this study are applicable to all oceans. The SST_S seems biased toward increasing trend in a time dependent, but unknown, manner; the results shown in Fig. 7 depending on the questionable assumption that the T_A are correct. Thus, SST, by itself, cannot be taken as representative of the oceans without correction. Furthermore, the ocean data sets thought to be free of SST instrumental bias effects appear to be as contaminated as those which admit a mix of measurement types.

There is hope that the oceanic T_A data, when thoroughly explored, will provide the needed ocean coverage. However, we can expect a legion of problems in this set that may be different than the SST_S problems (see the Appendix). Indeed, Höflich (1973) suggests the SST_S and T_A difference may be due largely to measurement changes associated with T_A and not SST_S. Hopefully these problems will not render the T_A data as suspect as the SST_S data.

As a final note, there is considerable interest in following the behavior of global or hemispheric temperature changes on an annual or pentad time scale. Ignoring for the moment that these quantities have not yet been estimated reliably, inspection of the error bars[3] (only ± 1-sigma, not 95% confidence bars) on Figs. 5 and 8 suggest that the grand averages are themselves highly uncertain. The discussion of significant changes in these averages on time scales less than decades must be of questionable usefulness. Such discussions, if they are to be responsible, seem to involve a burden of proof on the investigator that most likely cannot be met.

5. A modal change in the climate system?

Many of the regional and ocean averaged SST_S data definitely showed a relatively large "jump" (0.2°C) around the 1942 pentad. This feature was strongest in the North Pacific (Fig. 3b). Was the jump an actual geophysical happening associated with a transition from a pre-1940s mean climate state to a new post-1940s mean state as has been suggested by Joe Fletcher (personal communication, 1983)? He has reported that the HSSTD reflected an abrupt fall in both SST_S and T_A about 1900 and an abrupt rise about 1940. He states these changes are apparent in all ocean basins but occurred a few years earlier in the southern Indian Ocean than in other areas. He argues that these changes are probably real because they were accompanied by abrupt changes in the wind and atmospheric pressure fields and in the seasonal position of the ITCZ in the Asian sector.

On the other hand, could the discontinuities be the result of merging data sets with different properties? Several possible means of answering this question are discussed as follows. The reader should bear in mind the preliminary nature of the discussion due to lack of global data coverage and the fact that only temperature data are being considered here:

1) The hydrographic data from the specific North Pacific region are not numerous enough to investigate the early 1940s.

2) The T_A data for the North Pacific region (Fig. 3b) show the same discontinuity as the SST_S. However, the T_A do not suggest a change of mean state as clearly as do the SST_S.

3) The apparent discontinuity could be due to the merging of data from different sources. The individual data decks that were merged to make the Surface Marine Observations included the German and Netherland observations (decks 192 and 193) which effectively ended in 1938–39 (NCDC, 1968; Verploegh, 1967). The early British data (deck 194) cover 1850 to early 1950s (NCDC, 1968). These three sets together constitute the vast majority (17.5+ million observations) of the pre-1940 Surface Marine Observations. US Navy logs (decks 195, 110) and US Merchant Marine observations (deck 116) begin in early to late 1940s and constitute the majority of the data since that time. The European data are thought to contain large amounts of bucket measurements, while the latter sets are believed to be largely made up of injection temperatures (but no one knows for sure; see Section 3). Thus the possibility exists that the apparent early 1940s jump is largely the result of joining sets of bucket SST with sets of injection SST's. On a more global scale, the HSSTD decks are not supposed to have the "joining" problem. Yet the same shift is seen in the Indian Ocean. It is not seen in the Atlantic. However, earlier results (Section 3) suggest these data sets may well be subject to SST bias problems so their behavior in this case is of questionable value.

4) The yearly air temperature data over the three oceans were compared qualitatively with the SST_S series. The apparent discontinuity appeared in the 1940–42 period in most of the series; a result amply demonstrated in the ocean and all-oceans averages (Figs. 6 and 8). It was during this period that an apparently large El Niño event occurred and this may be what the SST_S/T_A data are reflecting. However, the jump does not seem associated with a change of mean state as suggested by the SST_S (e.g., Fig. 8).

[3] Other methods of estimating these error bars could change the conclusions that follow. Additional data, from a "full" ocean analysis, also could reduce the error bars.

5) The annual average air temperature for the Northern Hemisphere (Jones *et al.*, 1982) was checked for a similar discontinuity. Such a feature is not apparent in the pentad averages of that independent data set (see Fig. 8) nor in the annual values given by Jones *et al.* (cf. their Table 1).

6) The number of observations drops substantially during the early 1940s and there are some differences in the distribution of samples throughout the day (Appendix). However, the periods before and after the 1940s are relatively well sampled—certainly well enough to define pentad/oceanic averages. Thus it is not obvious how the space/time distribution of data could cause the apparent shift in mean.

It is concluded that the abrupt shift in SST_S may well be a partial result of merging data sets that contained different means of measuring ocean temperatures. The appearance of differing means of the SST_S for the pre- and post-1940s could well be explained by the resulting SST bias (see Fig. 7 also). All data, however, suggest a multi-year climate event did occur in the ocean–atmosphere system over much of the globe in the late 1930s through mid-1940s. The behavior of the independent T_A set substantiates this idea, as do separate analyses of sea level pressure in the central Pacific and average sea level along the west coast of North America (not shown). In a sense, then, an answer to the questions posed at the beginning of this section is "both possibilities appear partially true"; a rather unsatisfactory state of affairs. Hopefully, study of the event circa 1900 will be more satisfying.

A final note of interest is the apparent dichotomy between the results of this section and those of Section 4. The latter suggest a gradual increase in SST_S bias while the above results suggest a jump. The data from the Atlantic and Indian Ocean came from the HSSTD sets where a conscious effort was made to use ship measurements thought to be bucket only (i.e., European ships). The gradual instrument conversion on those ships is speculated to have led to the slow introduction of bias. The Pacific data used in this study are from a set that made no attempt to reject injection values and did join two sets of presumably different data types. Hence the jump effect which was present in nearly all Pacific averaging areas. Due to the large area covered by the Pacific and the areal weighting used to obtain Fig. 7, the jump would appear to occur in "all oceans."

The correctness of the above hypothesis can be tested by redoing the Pacific calculations for the HSSTD Pacific set (which was not readily available to us for this study) or by studying the distribution of data versus time and ocean by reporting country. The country code was not on the versions of the data available to us. In any event, the data shown in Fig. 7 should be regarded critically pending more exhaustive studies of the problems raised in this note.

6. Conclusions

Long time series of area averaged sea surface temperatures (SST_H) obtained from hydrographic data have been compared with series derived by identical procedures from ship observations in the Marine Deck (SST_S). The former data are noisy but do indicate a gradual measurement bias has entered the SST data presumably via the replacement of bucket with injection temperatures in modern times. In addition, the increase between 1899–70s in SST_S (~0.7°C) and the increase in ship reported air temperatures in the same averaging period/geographic regions (~0.4°C) differ by about the amount of the bias expected in going from bucket to injection temperatures. It is concluded that this apparent bias in SST data will make these data of questionable use in determining trends in ocean surface temperatures over the last century unless valid correction schemes can be developed. Use of the SST data in the trend determination will have to be rigorously justified with respect to the uniformity of measurement type. Simply stating that the data are "all buckets" or "all injection" will not be adequate since the data sets thought free of the bias effects (i.e., HSSTD sets) appear to demonstrate it as clearly as those decks known to contain the bias effect. Unfortunately, it may be very difficult to determine, for sure, the methods used to measure the SST, particularly in olden times.

The air temperatures over the oceans and the SST_S (in spite of their problems) both show an evolution of temperature since 1900 that differs in important aspects from surface temperature measurements over land. This suggests that reliable estimates of hemispheric and global temperature changes must adequately sample the variability over the oceans. Such reliable estimates have yet to be made.

The SST time series from Surface Marine Observations qualitatively show an apparent discontinuity of order 0.2°C in the period 1940–42 and is suggestive of a shift of climatic mean state. The T_A data from ships show a large climatological event beginning in the late 1930s, extending into the mid-1940s. However, a transition of mean states is not strongly suggested. The occurrence of an abrupt shift in SST means over most of the Northern Hemisphere ocean with no corresponding abrupt change in T_A means seems highly suspicious. Inspection of the individual decks that went into TDF11 suggest the apparent shift could be due to joining data sets that obtained their SST_S by different measurement methods. The approximate size of the qualitative discontinuity matches that expected from joining "mostly bucket" and "mostly injection" data sets.

In summary, the historical SST record from the Surface Marine Observations may well contain instrumental bias effects that render the data of questionable value in determining long period trends in ocean surface temperatures. The implications for using these data to estimate air–sea heat exchange may be even more dramatic (cf. Husby, 1980). Investigators that use the data in either of these quests, bear a heavy, perhaps impossible, responsibility for ensuring that the potential instrument bias has not contaminated their results. In any event, the relatively great uncertainty in large area/pentad averaged SST's will make it difficult to detect gradual changes in SST/T_A over the ocean that might be due to CO_2, volcanoes, etc. Even the assertaion that such changes are occurring seems difficult to make responsibly, without the aid of decadal averaging times and/or unusually large anomalies; a conclusion reached several years ago by Madden and Ramanathan (1980).

The reader is urged not to assume from this study that the historical set of ships' observations has so many problems as to be of questionable use for climate studies. Indeed, this set has taught us most of what we know about climate variations over the oceans. These same data sets hold rich promise of future scientific discovery, particularly when they are readily available to the full scientific community; a situation that should occur near the end of 1983 thanks to the efforts of Joe Fletcher and his colleagues at NOAA and NCAR. However, the user of these data must be aware that the sets do have limitations and that has been the point of emphasis. Hopefully more thorough study of the data now under way in the United States (Fletcher, personal communication, 1983) and United Kingdom (Folland, personal communication, 1983) will better define these problems and show ways to overcome them.

Acknowledgments. The author is indebted to Rob Quayle for many useful discussions, facts and good advice that made this paper possible. Comments on a preliminary draft by J. Fletcher, G. Paltridge, P. Jones, and K. Trenberth helped sharpen the text and are much appreciated.

This work was sponsored by the Lawrence Livermore Laboratories under contract to the Department of Energy via LLNL Purchase Order 9141701 and by the Climate Dynamics Office of the National Science Foundation under Contract ATM-8213279.

APPENDIX

Temporal Sampling Problems

It is logical to question whether some of the results obtained in this study could be obtained simply by changes in temporal sampling frequency. The answer to this question is explored in a preliminary manner in this Appendix. More complete analysis of the question seems warranted by the results presented below.

Perhaps the most crucial question for the measurement of air temperature and, to a lesser extent, sea surface temperature[4] are the distribution of observations throughout the day. The percentage of observations for each hour of the day is shown for a region in the central Equatorial Pacific by decade (Fig. A1). The histograms show that during recent decades the observations are relatively well distributed throughout the day. Indeed, in early times the observations occur more frequently during the day so even better temporal resolution of the diurnal cycle is realized. However, in the first two decades of the 1900s it appears that many of the regions (in the Pacific) had the majority of their observations taken at local noon (2100 GMT). These results are typical of the other Pacific regions. Without modification or appropriate weighting this will clearly lead to a bias in the estimates of daily and therefore yearly averaged air temperatures. Inspection

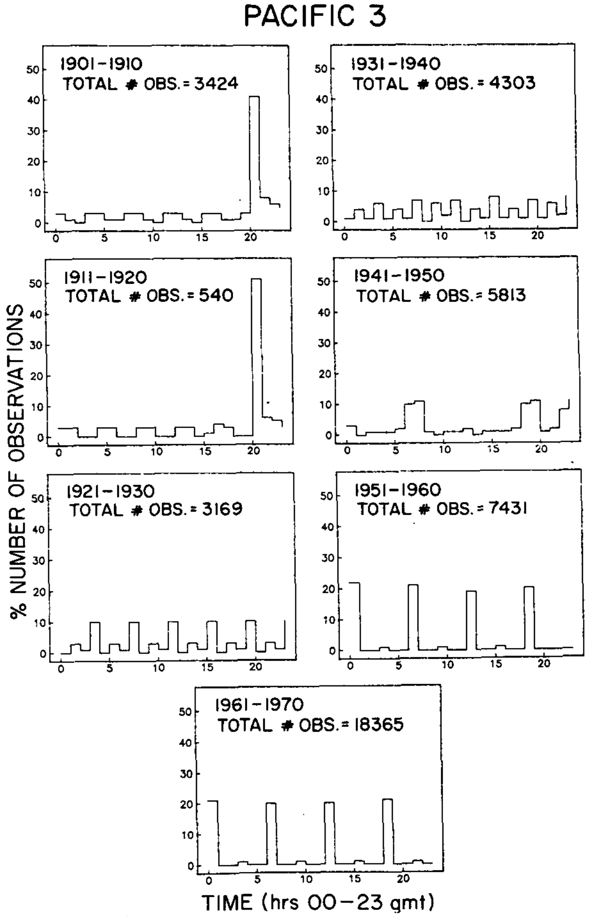

FIG. A1. Distribution of observations (percent) by hour of the day for each decade since 1900. The time is relative to Greenwich (GMT). The summaries are for Pacific Region 3 (see Fig. 2) which is in the central equatorial Pacific. In this area 2100 GMT corresponds approximately to local noon.

[4] Bucket measurements which come from the 'surface', will be much more susceptible to diurnal sampling problems than injection temperatures which are representative of greater depths.

of the instructions to observers indicates that prior to the early 1900s, the observations were to be taken at local noon (or more often if possible). A more exhaustive study of this particular sampling problem will presumably be undertaken by groups now preparing summaries of all existing marine observations.

A second temporal sampling problem concerns the number of observations as a function of time. The inserts in Fig. A1 again show the situation and hold no surprises. Since 1950 the number of observations has been relatively high. During World War II the observations drop in number by almost an order of magnitude. The decade of the 1930s was well sampled. However, prior to 1920s and 1930s the number of observations is roughly a factor of 10–30 less than in modern times. This general distribution of data seems typical of the Indian and Atlantic Oceans as well as the Pacific. Again more detailed analysis will hopefully be conducted with the more complete data sets now in preparation.

A final note of interest is the distribution of observations throughout the calendar year. Such a distribution is shown again for a typical region in the Pacific off Peru (Fig. A2). The results here are similar in many ways to those associated with the time-of-day observations. In the early years of the 1900s there appear to be clear sampling biases in the data towards specific months. However, once one comes to the middle and late 1900s, the data are quite well distributed throughout the year so that substantial biases in resolution of the annual cycle are not expected.

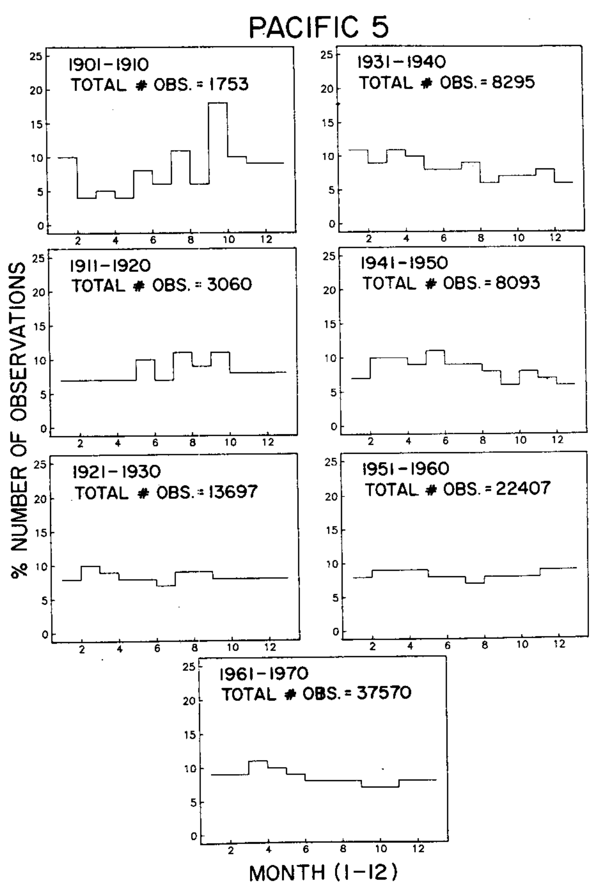

FIG. A2. Distribution of observations by month of the year for each decade since 1900. The summaries are for Pacific Region 5 which is the ocean region off the coast of Peru (see Fig. 2).

REFERENCES

Barnett, T. P., 1978: Estimating variability of surface air temperature in the Northern Hemisphere. *Mon. Wea. Rev.*, **106**, 1353–1367.

——, and J. D. Ott, 1976: Average features of the subsurface thermal field in the central Pacific. SIO Ref. Ser. 76-20, Scripps Institution of Oceanography, 84 pp.

Cayan, D. R., 1980: Large-scale relations between sea surface temperature and surface air temperature. *Mon. Wea. Rev.*, **108**, 1293–1301.

Folland, C., and F. Kates, 1982: Changes in decadally averaged sea surface temperature over the world 1861–1980. *Proc. Symp. on Milankovitch and Climate Change*, J. Imbrie and A. Berger, Eds., Elsevier (in press).

Goerss, J. S., and C. E. Duchon, 1980: Effect of ship heating on dry-bulb temperature measurements in GATE. *J. Phys. Oceanogr.*, **10**, 478–479.

Höflich, 1973: The seasonal and secular variations of the meteorological parameters on both sides of the ITCZ in the Atlantic Ocean. GATE Rep. No. 2, Part VI, World Meteorological Organization, 36 pp.

Husby, D. M., 1980: A comparison of surface heat flux estimates from Ocean Weather Station V and merchant vessels in its vicinity in the western North Pacific region, 1956–1970. *J. Phys. Oceanogr.*, **10**, 971–975.

James, R. W., and P. T. Fox, 1972: Comparative sea-surface temperature measurements. Rep. No. 5, Marine Science Affairs, WMO, No. 336, 27 pp.

Jones, P. D., T. M. L. Wigley and P. M. Kelly, 1982: Variations in surface air temperatures. Part I: Northern Hemisphere 1881–1980. *Mon. Wea. Rev.*, **110**, 59–70.

Madden, R. A., and V. Ramanathan, 1980: Detecting climate change due to increasing carbon dioxide. *Science*, **209**, 763–768.

Moyse-Bartlett, H., 1946: *From Sail to Steam: The Final Development and Passing of the Sailing Ship.* Historical Association of London, General Series No. 4, Staples Press, London, 20 pp.

National Climate Data Center, 1968: TDF11 Reference Manual, [available from NCDC, Ashville, NC].

Paltridge, G., and S. Woodruff, 1981: Changes in global surface temperature from 1880 to 1977 derived from historical records of sea surface temperature. *Mon. Wea. Rev.*, **109**, 2427–2434.

Reynolds, R. M., 1982: Comparison of surface meteorological observations from ship and toroid buoy in the North Pacific during STREX. *J. Appl. Meteor.*, **21**, 1032–1037.

Roll, H. U., 1965: *Physics of the Marine Atmosphere*, Academic Press, 30.

Saur, J. F. T., 1963: A study of the quality of sea water temperatures reported in logs of ships, weather observations. *J. Appl. Meteor.*, **2**, 417–425.

Seguin, W. R., and M. Garstang, 1971: A comparison of meteorological sensors used on the USCGSS *Discovered* during the 1968 Barbados experiment. *Bull. Amer. Meteor. Soc.*, **52**, 1071–1076.

Tabata, S., 1978: On the accuracy of sea surface temperatures and salinities observed in the northeast Pacific Ocean. *Atmos.-Ocean.*, **16**, 237–247.

van Loon, H. and J. Williams, 1976: The connection between trends of mean temperature and circulation at the surface. Part I: Winter. *Mon. Wea. Rev.*, **104**, 365–380.

Verploegh, G., 1967: Report of the consultant on the historical sea surface temperature program. WMO, Geneva.