Cloudy Winter Satellite Temperature Retrievals over the Extratropical Northern Hemisphere Oceans

NORMAN A. PHILLIPS

National Meteorological Center, Washington, DC 20233

11 March 1980 and 15 October 1980

ABSTRACT

An experimental colocation and statistical regression scheme is used to verify the hypothesis that large consistent mean errors in the cloudy oceanic satellite temperature retrievals north of 30°N from TIROS-N are due to a continental bias in the statistical colocation base. Sea surface water temperature is a useful predictor for these conditions. Significant improvement in these retrievals is possible. However, much of this improvement requires the presence in the colocation base of the mid-ocean radiosondes from the ocean weather ships.

1. Background

The National Environmental Satellite Service (NESS) obtains atmospheric temperatures (layer means) from its polar orbiting satellites by using linear regression methods to convert the measured infrared and microwave radiances to atmospheric temperatures (Phillips *et al.*, 1979; Smith *et al.*, 1979). The regression coefficients are revised weekly using a statistical data base of colocated satellite measurements and radiosondes accumulated for each of five latitude belts in the preceding several weeks. Most colocations north of 30°N are over continents. For example, only 40 maritime events were included in the 350 colocations in the belt 30–60°N used for the TIROS-N coefficient update by NESS on 11 November 1979. This meteorological unrepresentativeness for the important oceanic satellite data is most critical for cloudy retrievals, where heavy reliance must be placed on the microwave radiances since they are unaffected by nonprecipitating clouds. There are only four of these channels: MSU_1, MSU_2, MSU_3 and MSU_4, with respective peak responses at sea level, 600, 300 and 90 mb. MSU_1, furthermore, is seriously contaminated by its sensitivity to the capillary wave state of the sea surface.

In the past year, the NESS operational system, as evaluated at the National Meteorological Center (NMC), has resulted in the seasonal pattern of mean error (bias) of cloudy retrievals at oceanic stations between 32 and 66°N from TIROS-N that is shown in Table 1.[1] These are large enough to contribute significantly to the rms error. This bias pattern is weaker in the oceanic clear retrievals that can use the many tropospheric infrared channels. It is absent in the extratropical Southern Hemisphere maritime retrievals that use a much less continental colocation base chosen from those latitudes. In winter the cloudy retrievals can form as much as 65% of the oceanic data north of 30°N.

A crude physical interpretation of the bias pattern can be made if we note that the mean retrieved temperatures for the lower oceanic troposphere represent a lapse rate that is too stable in winter and too unstable in summer. This differential lapse rate is consistent with climatological data. The maps in Crutcher and Meserve (1970) give the following figures for area-averaged values of T (850 mb) minus T (500 mb) in the belt 30–60°N (value in °C):

	January	July
Continent	21.1	25.9
Ocean	22.9	22.6
Continental-ocean difference (C-O)	−1.8	+3.3

The climatological C-O differences here are consistent with the excessively stable cloudy oceanic retrievals produced by NESS in winter (and the too unstable retrievals produced in summer) from a continental colocation base. The hypotheses to be tested is that this bias pattern would be reduced with a strictly maritime colocation base.

2. Statistical procedures

A satellite colocation program developed at the National Meteorological Center was used to examine the effect of a maritime colocation base in winter. From 13 November 1979 through 20 January 1980, TIROS-N temperature retrievals and brightness temperatures (radiances, in effect) were colocated at 11 maritime locations: Bermuda, Azores, Sable Island, Valentia, Adak and Cold Bay in Alaska, plus the five meteorological ships. (NOAA-6 was not used because its MSU_3 became inoperative early in this period. The data set ended when TIROS-N stopped functioning temporarily on 20 January.) In addition to the NESS operational layer-mean retrieved temperatures, all four microwave (MSU) brightness temperatures (MSU_1, MSU_2, MSU_3, MSU_4) were colocated to the stations together with the four infrared channels ($HIRS_1$, $HIRS_2$, $HIRS_3$, and $HIRS_{17}$) that peak in the stratosphere and are therefore less sensitive to clouds (Smith *et al.*, 1979). The value of the sea surface water temperature (SST) at the station was also recorded from the NMC analyzed fields. The colocation process consisted of collecting all TIROS-N reports from one orbital pass within 600 km of a station and then, subject to the condition that the station must be surrounded in all four quadrants by these retrievals, mapping the reports to the station location (Phillips *et al.*, 1979). Colocations in which the interpolated microwave radiances were likely to have been seriously contaminated by rain ($MSU_2 - MSU_1 > 12$°C) were discarded.[2] Successive 12 h radiosonde temperatures were then interpolated to the time of the satellite passage. Only radiosondes satisfying the most restrictive NMC data processing checks were used, and all disagreements greater than 10°C between radiosonde and retrieved temperature were individually inspected for possible radiosonde errors. None were found. A total of 981 colocations were achieved, a collection rate of slightly more than one per station per day per satellite.

Each original retrieval entering into a colocation event was assigned a number e = 300, 200 or 100 according to whether it was a clear column, partly cloudy or cloudy retrieval, corresponding to the first (A), second (B), or third (C) retrieval

TABLE 1. Seasonal pattern of differences (satellite minus radiosondes) in mean layer virtual temperatures (°C) at cloudy oceanic stations.

Layer (mb)	Feb 1979	Aug 1979
1000–850	−2.1	+4.4
850–700	1.5	1.1
700–500	2.1	−0.0
500–400	2.0	−0.7
400–300	1.6	−1.2
300–200	1.3	1.1
200–100	−0.3	0.2

---

[1] The word error as used in this paper really means "difference from radiosonde values." When these errors become smaller than 1°C, this usage must stop.

[2] This oceanic precipitation test was introduced operationally by NESS in mid-February 1980.

TABLE 2. Division of the colocation sample into five periods.

Period	Length (days)	Number of colocations $e < 150$	Number of colocations $e > 200$
13 Nov–30 Nov	18	95	53
1 Dec–14 Dec	14	97	68
15 Dec–26 Dec	12	81	40
27 Dec– 7 Jan	12	67	45
8 Jan–20 Jan	13	90	47

paths in NESS terminology (Phillips *et al.*, 1979). These individual e values were also mapped to the radiosonde location, so as to allow organization of the data by retrieval paths. The resulting colocation events were

$e > 200$ (essentially clear column retrievals)	253
$150 < e < 200$ (indeterminate)	298
$e < 150$ (essentially cloudy retrievals)	430
Total	981

Cloudy retrievals are less accurate than clear column retrievals (Phillips *et al.*, 1979) and the large number of $e < 150$ colocations, reflecting the high cloudiness found over the extratropical winter oceans, demonstrates the importance of improving their accuracy.

The colocation data thus gathered consisted of three types of information:

(i) NESS operational retrieval temperatures (for seven layers)

(ii) Radiosonde temperatures (for seven layers) as predictands

(iii) Eight NESS operational brightness temperatures plus sea surface water temperature, forming nine potential predictors for (ii).

In an accepted colocation, types (i) and (iii) were always complete, but some radiosonde mean layer temperatures could be missing because of the NMC processing checks, a failure to reach 100 mb, and a requirement that the 1000–850 mb temperature was accepted only if the surface pressure was >985 mb.

The data set having $e < 150$, for the 32-day period 13 November–14 December, was first examined for me by Frank Lewis with a statistical regression screening procedure. This process, separately for each of the seven predictands, determined the sequential importance of each predictor in explaining the variance in that predictand after allowing previously selected predictors to operate. I then divided the entire $e < 150$ data set into the five consecutive periods shown in Table 2. The following treatment was applied to the $e < 150$ data, for each predictand separately. A sequence of seven computations was made, using first the three most helpful predictors for that predictand uncovered in the screening process, then adding the fourth most helpful, etc., ending up with all nine predictors in the seventh computation. In each of the seven computations for one predictand, five sets of regression coefficients were determined by treating each of the five periods as a new data set. In each of the last four periods, coefficients determined in the previous period also were used to predict the predictand from current predictor values, and these predictions were verified against the colocated radiosonde values. This imitates the NESS coefficient update procedure, and tests the regression on independent data. The seven values of total rms error for the last four periods thus obtained for each predictand were accepted as a measure of the success for that predictand, using those seven sets of predictors (three to nine in number).

As the number of predictors was increased from 3 to 9 for a given predictand, this error reached a minimum, but then increased as the last predictors were called into play but did not hold up on independent data. The minimum rms error point for each predictand identified its most reliable set of statistical predictors. These choices are shown in Table 3.

Cumulative error measures for the NESS operational retrievals also were accumulated over the last four periods, not only for the identical $e < 150$ colocation events treated above, but also for the different set of 200 clear retrieval colocation events characterized by $e > 200$. The latter provide a reference against which the two different cloudy retrievals—NESS and experimental—can be judged.

Before examining the performance of the new statistical system for third path (cloudy) retrievals, I summarize the principal differences in procedure (some of which have not been emphasized above).

1) The new method is more aggressive in finding colocations, as evidenced by its high return rate of more than one colocation per day per station per satellite.

2) In the new method, the coefficients are determined from only marine colocations.

3) In the tests reported in the next section, the skill for both systems was measured at the same stations used to derive coefficients in the new method.

4) The new method used sea surface water temperature as an additional predictor.

5) The new method has a more direct statistical regression between predictors and the seven-layer temperatures than does the NESS operational system. The latter uses empirical orthogonal functions and an intermediate set of temperatures at 40 pressure levels.

6) Coefficients in the new method are obtained

from a much smaller colocation base (~90) than is used for the NESS system (~400).

7) Coefficients for these cloudy retrievals in the new system are based on primarily cloudy colocations, whereas the NESS colocation base is not only mostly continental, but is restricted to clear or partly cloudy events.

3. Results

An ideal test of the continental-ocean bias hypothesis would have been to parallel simultaneously the experimental procedure described above with a second one in which the experimental regression coefficients were determined from only continental colocations. Seawater temperature could be omitted as a predictor for both procedures, and both experimental results would be evaluated against an independent set of oceanic verification radiosondes. The latter stations do not exist, and resources did not readily permit the parallel continental colocation base. A critical point for the hypothesis therefore is the extent to which the operational cloudy retrievals reproduce the February 1979 bias figures given in Table 1, and the extent to which the experimental retrievals reduce that bias.[3]

Fig. 1 shows this bias data, and verifies the hypothesis. The NESS cloudy retrievals from TIROS-N again show their Northern Hemisphere winter ocean bias pattern, although not quite as extreme as those for February 1979. The experimental values correct this, especially when allowance is made for the estimated sampling error of the mean difference (bias) that is indicated on the left side of the figure as about 0.2°C in magnitude, and derived from the relation

$$\sigma\ (\text{mean}) \approx \frac{\text{standard deviation}}{(\text{sample size minus one})^{1/2}}\ .$$

The experimental cloudy bias is also less than the NESS clear biases. This is gratifying, but it should be kept in mind that the station sampling varies from the cloudy ($e < 150$) set to the clear ($c > 200$) set. In particular, Bermuda and the Azores, the two most southerly stations, contributed only 9% of the $e < 150$ colocations, but 31% of the $e > 200$ colocations. The $e > 200$ set had radiosonde temperatures suggestive of a sharper tropopause than that experienced in the $e < 150$ data set. This would present a more difficult retrieval problem for any remote sensing system.

Fig. 2 shows the rms errors. The new system is an improvement over the operational NESS cloudy retrievals in the four lower layers.[4] Both $e < 150$ error curves reflect the 600 mb peak response of MSU channel 2. The original screening run by Frank Lewis on the $e < 150$ colocations for the period 13 November–15 December determined the percentage of variance explained by each predictor as it was called into action. MSU channel 2 was the first selected for the bottom five layers (Table 3). In these five layers it explained, respectively, 84, 90, 93, 90 and 86% of the variance. The resemblance of this purely statistical result to what would be expected from this sensor on purely physical grounds suggests that the larger errors in cloudy retrievals in the layer 500–300 mb could be reduced by an additional microwave channel with peak response at 400 mb. An additional channel with a peak around 850 mb might also be useful.

A final interesting fact shown on Fig. 2 is the pinpointing of the 700–500 mb layer as the layer for which the experimental system makes the greatest improvement over the operational cloudy retrievals. This suggests to me that the numerical aspects of the NESS operational statistical regression system (quite apart from its poor choice of colocation locations) may not be taking full advantage of the information contained in MSU channel 2.

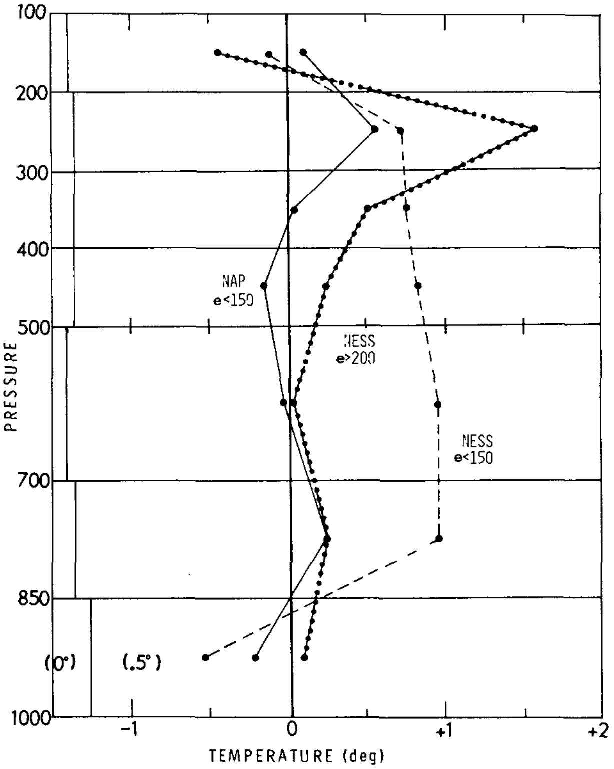

FIG. 1. Bias (mean error) for cloudy ($e < 150$) and clear ($e > 200$) TIROS-N temperature retrievals at 11 oceanic stations, 32–66°N, 1 December–20 January. The left curve, with its own horizontal scale, approximates the sampling error of the mean difference for cloudy retrievals.

---

[3] The NESS Aug. 1979 bias pattern was repeated in the summer of 1980 (with the NOAA-6 satellite).

[4] Much of the reduced rms is due to reduced bias, especially in layers 2, 5 and 7. This, of course, was the main hypothesis of the test.

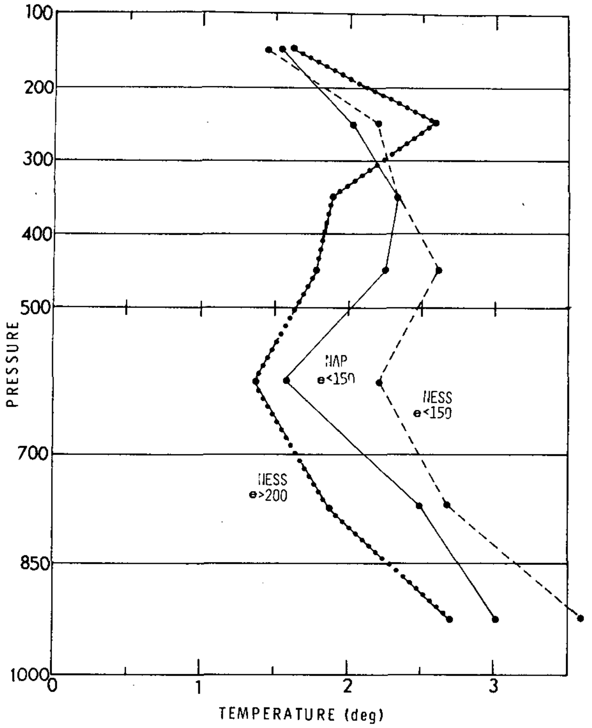

FIG. 2. Root-mean-square temperature error for cloudy ($e < 150$) and clear ($e > 200$) TIROS-N retrievals at 11 oceanic stations, 32–66°N, 2 December–20 January.

An important use of satellite temperatures is to construct hydrostatic fields of isobaric height

$$h(p_{n+1}) = h(1000) + \frac{R}{g}\sum_{j=1}^{n} \bar{T}_j \ln(p_j/p_{j+1}),$$

where the sum is over all layers $p_{\text{bot}} > p > p_{\text{top}}$ below the pressure surface p, and $h(1000)$ is an independently obtained height of the 1000 mb pressure surface. Fig. 3 shows the results obtained by computing h as shown above, but with $h(1000) \equiv 0$, for retrieved temperatures and for radiosondes. (The sample size in this computation must decrease monotonically with decreasing p, since the first missing radiosonde temperature in a colocation event prevents computation of h errors above that layer.)

The experimental $e < 150$ results reduce the NESS $e < 150$ height errors to a value not much larger than the NESS clear column errors. All three methods on Fig. 3 have rms height errors that are smaller than would exist if the temperature errors in any one retrieval were always of one sign. This can even be seen qualitatively, by using the reference radiosonde height error curve on Fig. 3, which has been computed assuming a uniform temperature error of +0.5° in all layers, to convert any vertically averaged rms value taken from Fig. 2 to an equivalent height error. This beneficial effect on height errors is a practical expression of the retrieval temperature error correlations between different levels that have previously been evaluated for NESS remote soundings (Phillips *et al.*, 1979; Schlatter and Branstator, 1979). It is important that the new cloudy retrieval results have a pattern of these correlations which is at least as favorable as that for the operational NESS cloudy retrievals when isobaric heights are computed.

Experience with the retrievals from the current TIROS-N satellite (and with its predecessor experimental version on Nimbus 6) has shown a common tendency to underestimate the coldest or warmest temperature extremes. This implies that the retrieval errors are negatively correlated with the true field. Fig. 4 presents these correlations for the two $e < 150$ retrieval systems. Both systems exhibit this negative correlation. Between 200 and 700 mb, however, the new system is evidently more willing to predict extremes, and one can therefore expect the reduced rms errors from it in this region to be very useful in capturing temperature extremes in cloudy regions. At 925 mb the new system has a larger negative correlation than does the operational scheme, however, so that the reduced rms error of the new system at 925 mb (Fig. 2) may not, by itself, result in an immediate improvement in analysis of

TABLE 3. Choice of predictors for each layer-mean temperature. Numbers denote the sequential importance uncovered in the screening procedure.

	Predictor and its peak sensitivity level (mb)								
Layer (mb) (predictand)	SST (1000)	MSU_1 1000	MSU_2 600	MSU_3 300	MSU_4 90	H_1 30	H_2 60	H_3 100	H_{17} 2
1000–850	2	3	1						
850–700		4	1	2	3		5		
700–500	3	5	1	2	4		6		
500–400	3		1	2		4			
400–300	4		1	5	2				3
300–200	4		3	1	2				5
200–100	6		5	4	1		2	3	

extreme temperatures at this level. Fortunately, over the oceans there is auxiliary temperature information at sea level in the form of ship *air* temperatures that can improve analyses at low levels in the Northern Hemisphere.[5]

The role of the sea surface temperature as a predictor was examined by omitting it as a predictor in layers 1 and 3–7 (see Table 3) and adding replacement predictors to some layers: MSU_3 to layer 1 and MSU_1 to layers 4–7. The NAP rms values on Fig. 2 were increased by this change: layer 1 (+0.49°C), layer 2 (no change), layer 3 (+0.19°C), layer 4 (+0.27°C), layer 5 (+0.10°C), layer 6 (+0.35°C), layer 7 (+0.00°C). The sea surface temperature is therefore a very useful predictor in this winter data set. It is perhaps worth pointing out that it can only be used as part of a purely maritime colocation base.

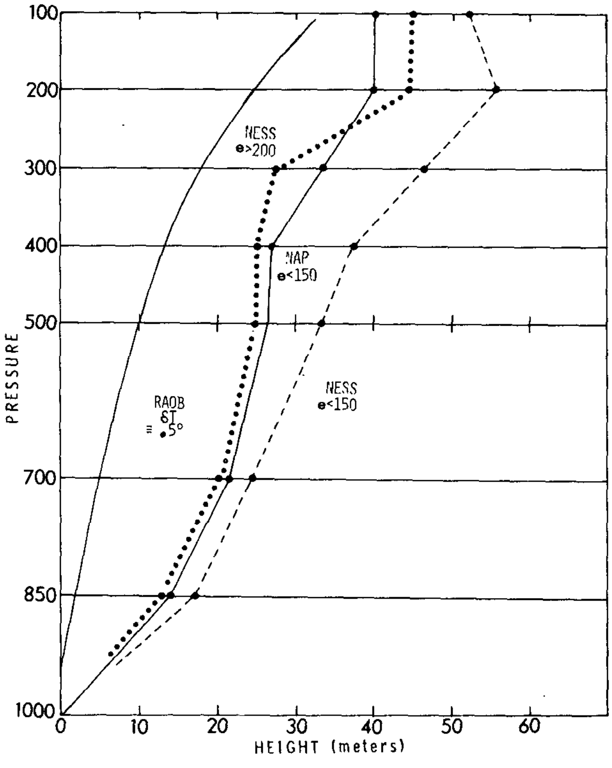

FIG. 3. Root-mean-square isobaric height error relative to 1000 mb for cloudy ($e < 150$) and clear ($c > 200$) TIROS-N retrievals at 11 oceanic stations, 32–66°N, 1 December–20 January. The left curve is for a radiosonde with a constant 0.5° temperature error.

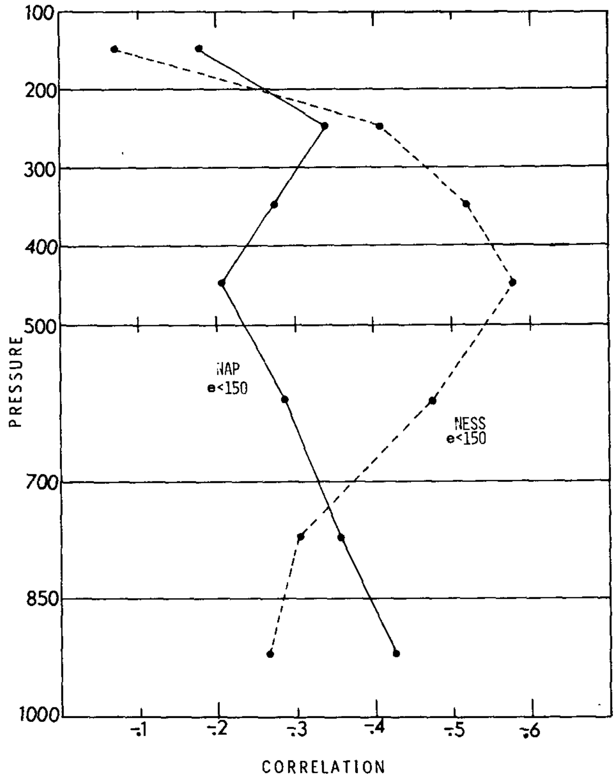

FIG. 4. Averaged correlation coefficient between retrieval error and true temperature for cloudy TIROS-N retrievals at 11 oceanic stations, 32–66°N, 1 December–20 January.

4. The importance of the ship radiosondes

The preceding section has demonstrated that cloudy winter oceanic retrievals made from a maritime colocation base can be better than those from the present NESS continental colocation base. The role of the ships as part of this maritime base can be assessed to some extent from the available data base of 430 cloudy colocations, although this is limited in size. To cope with this limitation, the $e < 150$ data was divided into the three (instead of five) periods of almost equal total colocation numbers as shown in Table 4.

The evaluation program was run twice on these three time periods. In both runs, verification was performed at all 11 stations, as giving the best overall measure of success for the oceans. In both runs, the experimental method again used coefficients determined from the previous period (i.e., these tests are on independent data for the period 8 December–20 January). In one run the experimental regression coefficients were obtained using all 11 stations. This run differs from those de-

---

[5] In dependent samples, this correlation coefficient equals the negative square root of the ratio (error variance/true variance), since the error is uncorrelated with the predicted value in a dependent sample. In each of the four independent periods, the error variance for the NAP method in the layer 1000–850 mb was less than that for NESS, and, if the dependent relation for this correlation coefficient had been valid, would have produced an average correlation coefficient of −0.47 against a value of −0.56 for NESS. This layer had only 175 colocations in the last four periods compared to almost twice as many in the other six layers.

TABLE 4. Sample design for testing the role of ship radiosondes.

Period	Days	$e < 150$ colocations 5 ships	 6 islands	Total
13 Nov–7 Dec	25	68	73	141
8 Dec–28 Dec	21	76	67	143
29 Dec–20 Jan	23	88	58	146

scribed in Section 3 only in its use of three instead of five time periods. In the second run, however, the experimental regression coefficients were obtained using only the six island stations.

Results are shown in Table 5. The NESS results and the experimental results using all 11 stations are essentially identical to those shown in Fig. 2 for the period 1 December–20 January. The experimental results from the island colocation base are (except at 200–100 mb) worse than those from the complete maritime base. The last column headed x in the table shows the following ratio of rms values:

$$x = \frac{\text{NESS} - \text{Islands}}{\text{NESS} - \text{All}} .$$

It expresses the extent to which the improvement over the NESS results achieved by the new colocation system at all 11 station locations would be retained if the ship radiosondes were not available in the coefficient base. The loss in improvement is appreciable at all layers except 1000–850 and 200–100 mb. Since results for these two layers are based on the two smallest sample sizes, their anomalous character could be a sampling accident. An alternate interpretation for 1000–850 mb is that its three predictors (SST, MSU_1, MSU_2) are capable of handling considerable differences between the dependent and independent data sets.

The mean radiosonde temperatures at (i) the 11 verification stations for the verification period (8 December–20 January) did differ significantly from (ii) those at the six island stations for the period when coefficients were determined in the above test (13 November–28 December). The former mean temperatures were colder by 2.5–3.0° in the six bottom layers, while for the 200–100 mb layer the difference was only +0.29. (The latter may explain the anomalous result in Table 5.) Whether this difference is due to time changes or to regional sampling differences cannot be addressed meaningfully with the limited available data base.

The reader may wonder why additional coastal stations (or Iceland, for example) were not included in the non-ship part of the maritime co-locations in order to obtain a better statistical base. Meteorological conditions at Vancouver in Canada, for example, are not that different from those at the fixed ship PAPA 1200 km or so to the west. The problem here is not one of meteorological representativeness but rather the unpleasant physical facts that the brightness temperature of microwave channel 1 increases suddenly by 30–40°C from ocean to land because of different surface emissivities, and coastal mountains will reduce the brightness temperature seen by the important second microwave channel, just as cumulus clouds affect infrared radiances. These very sharp geographical effects cannot be disentangled in the radiance retrieval data available from NESS, since the latter are averages over retrieval "boxes" of typical size $(250\ \text{km})^2$. It is doubtful if these effects can even be allowed for in the original spot microwave radiances, since these sense a circle of 110 km or so in diameter (Smith *et al.*, 1979). Not only is this distance large enough to introduce random noise into colocation statistics by limiting the smallness of the horizontal distance between satellite and radiosonde locations, but an extremely detailed representation of individual mountains would have to be available to the NESS radiance processing system to identify effects of individual mountains.

The paucity of radiosondes on small islands in these latitudes therefore assigns a major role to the ocean ship radiosondes in providing a realistic data base for cloudy retrievals. When this expensive ship data is exploited fully, however, the satellite system can act as a powerful multiplier of that data by providing temperature retrievals of reasonable accuracy in the large areas between the ships, even in cloudy areas.

Dr. C. M. Hayden, in refereeing this paper prior to publication, argued that I may have overestimated the need for an active ship radiosonde program, because microwave retrieval coefficients could also be obtained from a colocation base formed from

TABLE 5. Root-mean-square differences between cloudy retrieved satellite temperatures and 11 maritime radiosonde stations for the period 8 December–20 January. Results are shown for NESS operational retrievals and for experimental retrievals using both 11 stations and only the 6 island stations for regression coefficients. x is explained in the text.

Layer (mb)	Sample	NESS	All	Islands	x
1000–850	153	3.69	2.86	2.92	0.93
850–700	283	2.70	2.39	2.46	0.77
700–500	280	2.27	1.60	1.84	0.64
500–400	272	2.66	2.25	2.45	0.51
400–300	266	2.38	2.31	2.51	−1.86
300–200	267	2.24	2.15	2.43	−2.11
200–100	238	1.49	1.64	1.59	**