Kent, E. C., P. K. Taylor, B. S. Truscott, J. S. Hopkins, 1993: The Accuracy of Voluntary Observing Ships' Meteorological Observations-Results of the VSOP-NA. *J. Atmos. Oceanic Technol.*, **10,** 591–608. doi: 10.1175/1520-0426(1993)010<0591:TAOVOS>2.0.CO;2

The Accuracy of Voluntary Observing Ships' Meteorological Observations—Results of the VSOP-NA

ELIZABETH C. KENT AND PETER K. TAYLOR

James Rennell Centre for Ocean Circulation, Chilworth, Southampton, United Kingdom

BRUCE S. TRUSCOTT AND JOHN S. HOPKINS

Meteorological Office, Bracknell, Berkshire, United Kingdom

(Manuscript received 17 November 1992, in final form 24 November 1992)

ABSTRACT

For the Voluntary Observing Ships Special Observing Project for the North Atlantic (VSOP-NA), the layout, meteorological instrumentation, and observing practices of 45 voluntary observing ships (VOS) operating in the North Atlantic were cataloged. Over a two-year period these ships provided extra information with each observation, and the effect of different observing practices has been quantified by using analysis fields from an atmospheric forecast model as a comparison standard. Biases of order several tenths of a degree Celsius were detected in sea surface temperature data from engine intake thermometers, in dewpoint temperatures from screens (and to a lesser extent, psychrometers), and in air temperatures due to solar heating. Wind speeds from anemometers were high compared to visual winds by about 2 kt for winds up to about 25 kt. The VSOP-NA data do not, however, indicate which is the more accurate. Correction for anemometer height and use of the WMO Commission for Marine Meteorology version of the Beaufort scale reduced this difference significantly. The result of these corrections on mean heat flux estimates was only a few watts per square meter but much greater changes resulted for particular areas and seasons. The project identified observing methods that are to be preferred for future use on the VOS, and demonstrated that the combined use of VOS data and a forecasting model allowed the detection of biases both in the observations and in the model analyses.

1. Introduction—Errors in VOS observations

The atlases of the marine climate compiled by Bunker (1976), Esbensen and Kushnir (1981), Hellerman and Rosenstein (1983), Isemer and Hasse (1985, 1987), Oberhuber (1988), Wright (1988), and others are all based on the meteorological observations from the voluntary observing ships (VOS). These data have been assembled and made available in such datasets as COADS [the Comprehensive Ocean–Atmosphere Data Set, (Woodruff et al. 1987)]. However, it has long been recognized that the observations made on board these merchant vessels are subject to both systematic and random errors. The observations are performed on a voluntary basis by the ships' officers, different recruiting countries favor different measurement techniques, and it is difficult to site instruments so as to provide measurements representative of the air undisturbed by the ship. While many of the errors are likely to be insignificant for operational forecasting purposes, even small biases can have significant effects in air–sea flux estimations (Blanc 1986). For example, in estimating possible climate changes, daytime marine air temperature observations have been ignored because of errors due to solar heating (Jones et al. 1988), and corrections for different measurement methods are required to produce a systematic sea surface temperature dataset (Bottomley et al. 1990). Some national services consider that Beaufort scale estimates of wind speed are preferable to measured values that are subject to unquantified calibration, exposure, and reading errors. However, different versions of the Beaufort scale result in different climatologies (Isemer and Hasse 1991), and the increasing fraction of ships using anemometers can result in a spurious climatic trend (Cardone et al. 1990).

Unfortunately, while it is recognized that the measurement method can bias the VOS observations, the data submitted by the VOS both in real time and later via log sheets include very limited information to allow discrimination between different observing practices. Furthermore the instrumentation information provided in the list of VOS (WMO 1990) is lacking in detail and in some cases is inaccurate. For this reason the Voluntary Observing Ships Special Observing Project for the North Atlantic (VSOP-NA) was set up to establish the effect on VOS data of different instrumentation and observing practices. This paper will

Corresponding author address: Elizabeth C. Kent, James Rennell Centre for Ocean Circulation, Gamma House, Chilworth Research Centre, Chilworth, Southampton SO1 7NS, UK.

summarize the VSOP-NA results (Kent et al. 1991). It will first discuss the characteristics of the VSOP-NA ships and the method of analysis of the VSOP-NA dataset (section 2). The results, showing the effect of the different measurement methods, will be presented in section 3 and will be discussed in the context of previous investigations in section 4.

2. Description of the VSOP-NA project

a. *The VSOP-NA ships*

An initial selection of suitable ships was made on the basis of the reporting performance of the VOS regularly plying the North Atlantic. Recruitment was undertaken by port meteorological officers who obtained information on instrumentation and observing practices. Forty-five ships were selected from the national observing fleets of Canada, France, Germany, the Netherlands, the United Kingdom, and the United States. The instrumentation carried by these ships was documented by the port meteorological officers and a catalog of the VSOP-NA ships (Kent and Taylor 1991) was prepared containing, for most vessels: the regions from which reports were received, the method of observation of wind, air temperature, humidity, sea surface temperature, and pressure; the height of all instruments above a nominated reference level; ships' plans clearly showing the instrument positions and instrument exposure ratings; and limited model–manufacturer details for the meteorological instruments in use.

The typical VSOP-NA ship was a container vessel of about 210 m in length that traveled at 17 kt (8.5 m s^{-1}) (Figs. 1a and 1b). It was loaded with cargo to about 10–20 m above the main deck. Sea temperatures were measured by bucket, or by engine intake, or hull contact sensors at depths between 3 and 9 m (Fig. 1c). The air temperature and humidity observations were taken about 20–30 m above sea level, and the anemometer, if carried, was at about 30–35 m (Figs. 1d and 1e). Comparison with the anemometer heights for the VOS fleet as a whole (Fig. 1e) suggests that the VSOP-NA ships were probably biased toward greater length and therefore higher observing platforms.

In general the mix of instrument types used on the VSOP-NA ships was similar to that for the VOS fleet as a whole (Table 1). This was true for the fraction of visual and anemometer wind estimates (about two-thirds are visual), and for screen and psychrometer temperature and humidity measurements (roughly half and half with the VOS biased toward psychrometers and the VSOP-NA toward screens). For sea surface temperature measurement, 52% of the VSOP-NA ships used buckets, compared to 32% of the VOS. There were also more VSOP-NA ships with hull contact sensors. A higher percentage of the VSOP-NA ships used digital aneroid barometers to measure air pressure rather than the analog aneroid barometers used by most of the VOS fleet.

Despite the various differences noted above, it is considered that the observations from the VSOP-NA ships are typical of the bulk of observations from the North Atlantic and therefore a good indicator for the accuracy of VOS reports from that area.

b. *Assembly of the dataset*

The VSOP-NA ships were requested to provide extra information with each report. Entered on special VSOP-NA log sheets, this supplementary information consisted of the ship's speed and heading; height of deck cargo above sea level; height of nominated reference level above sea level; method of sea surface temperature measurement; location of air temperature measurement; and the relative wind speed and direction. The VSOP-NA log sheets were collected by the port meteorological officers and sent to the Deutscher Wetterdienst Seewetteramt for transferring to magnetic tape, and hence to the U.K. Meteorological Office (UKMO) where the VSOP-NA ship observations were matched with the corresponding values from an atmospheric forecast model (see section 2c). None of the data, except reports from the German ships, were quality controlled. The dataset will therefore contain errors comparable to those present in real-time data from the Global Telecommunications System (GTS) of the World Weather Watch. The small number of observations containing errors in the ship position were, however, deleted. Such errors may have occurred while coding or archiving and would normally be removed at a forecasting center by standard quality control procedures.

The dataset of matched observations was then sent to the James Rennell Centre for Ocean Circulation (JRC) for analysis. Codes for the different types of instrumentation used on the ships were defined and included with each observation in the dataset. To allow a quantitative analysis of the exposure of the various sensors, an exposure index was devised. Ratings depended on the distance to an upwind obstruction, whether the flow was fully or partially blocked by that obstruction, and, for well-exposed sensors, the upwind fetch over the ship (Table 2). For each sensor the index was estimated for relative wind from each of four quadrants centered on the bow, beam, and stern of the ship. The appropriate exposure rating was then associated with each observation depending on the reported relative wind direction.

The total number of observations archived in the observation dataset was 33 736. In addition, 3656 observations from OWS *Cumulus,* when on station LIMA (57°N, 20°W), were collected and added to the archive. Of these 37 492 observations, 11 215 observations were not included in the matched comparison dataset used for the analysis. Most of these ship reports

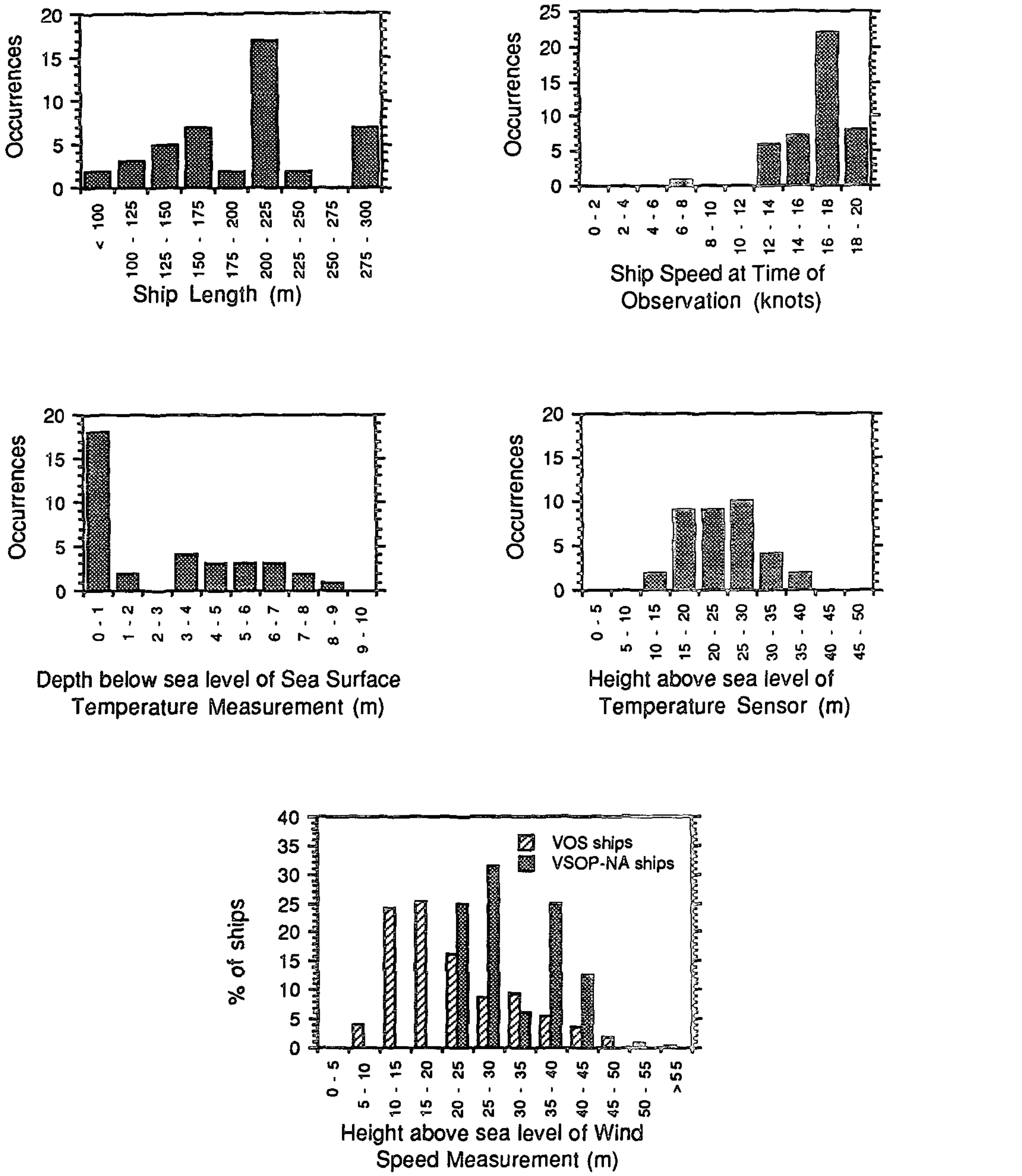

FIG. 1. Characteristics of the VSOP-NA ships: (a) ship length; (b) ship speed at the time of observation; (c) depth of sea surface temperature measurement (bucket values are shown as 0–1 m); (d) height of the temperature and humidity measurement; (e) height of anemometer (values for the whole VOS fleet are also shown).

omitted from the analysis were outside the area of the atmospheric forecast model. This left 26 277 observations archived along with model values and special VSOP-NA information in the comparison dataset. Geographical filtering was then applied to remove data close to some areas of land as it was felt that the model could not adequately represent local coastal effects, especially around Iberia and the east coast of North America. Thus the dataset on which the results presented here are based contained 24 952 observations. Figure 2a shows the geographical distribution of these data and a calendar of matched observations is shown in Fig. 2b country by country. The data collection started in spring 1988 and continued until summer 1990. A full annual cycle of observations was achieved by 28 of the 45 vessels, with some supplying two years of data. In contrast, eight vessels submitted data over a period of six months or less.

TABLE 1. Percentages of VSOP-NA and VOS equipped with particular meteorological instrument types.

Parameter	Method of measurement	Percentage of fleet VSOP-NA	VOS
SST	Bucket	52	32
	Engine-room intake	38	65
	Hull contact sensor	4	2
	Other	—	1
	Unknown	8	—
Temperature and humidity	Screen	49	44
	Psychrometer	38	55
	Unscreened	—	1
	Unknown	13	—
Wind speed and direction	Fixed anemometer	17	22
	Hand-held anemometer	7	8
	Visual	63	70
	Unknown	13	—
Pressure	Precision aneroid barometer	40	9
	Analog barometer	60	89
	Mercury barometer	0	2

c. The model as a comparison standard

To provide a reference to allow observations to be compared, each VSOP-NA observation was matched against a corresponding numerical weather prediction model analysis value. The UKMO Limited-Area (so-called fine mesh) Model (Alves 1991, 1992; Bell and Dickenson 1987) was selected as a comparison standard for the VSOP-NA observations because, of the models operating at the UKMO at the time, this model was expected to lead to the smallest interpolation errors in both space and time. The fine-mesh model operated on a grid resolution of 0.75° latitude × 0.9375° longitude over an area 30°–80°N, 80°W–40°E, and eight analyses were archived daily.

The fine-mesh model used VOS observations extracted from the UKMO Synoptic Data Bank, which contains all data received via the GTS. As part of the operational routine, these data are quality controlled by checking against background field values and other observations in the area. The observations are then incorporated into the analysis using a data assimilation scheme which takes into account the type of observation and the observation time relative to the analysis. A high weighting indicates a reliable observation taken close to the time of the analysis: such an observation would have a large influence on the analysis. The weighting can be set to zero so any particular observation type can be ignored by the analysis if required. Bogus observations (created by forecasters in real time and based on qualitative evidence, such as satellite pictures) may have been added to the analysis with a high weighting if it was considered that the model did not represent a feature well. The archived analysis fields may thus contain an element of subjectively introduced information.

The fields of interest to the VSOP-NA are: sea surface temperature (SST), 10-m wind speed and direction, mean sea level (MSL) pressure, and 1.5-m air temperature and relative humidity. The model humidity values archived in the VSOP-NA dataset are dewpoint temperatures calculated from the model relative humidity and dry-bulb temperature. The model analysis scheme combines all valid observations with a background field to produce a "best estimate" of conditions on the synoptic scale. For the atmospheric variables, the background field is an earlier forecast field and the analyses are updated at 3-h intervals. For SST, the analysis scheme is based on observations from seven different data sources, a climatological field of SST, and a background field (essentially persistence). The data are extracted from the UKMO Synoptic Data Bank at approximately 1200 UTC each day for observations between 0001 and 2400 UTC for the previous day. After a series of validity checks, the SST at a grid point is calculated by combining climate and background fields with valid observations within a defined immediate area. Ocean weathership (OWS) reports have a weighting of 4.0, fixed buoys and expendable bathythermographs have a weighting of 1.5 and VOS, drifting buoys, and satellite observations have a weighting of 1.0. Progressively more weight is given to the climate field and less to the background field as the time interval between observations and the previous analysis increases.

The model data assimilation scheme does not make use of all observations in the analysis of the various fields. For the VOS and OWS reports, air temperature and humidity are not used, and only the pressure, wind, and sea surface temperature are assimilated into the model. For those variables, the VSOP-NA observation may therefore have influenced the model value it is being matched with. This problem could have been removed by comparing the observation with a forecast field, but most of the variables required were not available in the forecast archive. The model analysis was therefore used as a comparison standard and the ap-

TABLE 2. Exposure definitions for VSOP-NA sensors.

Rating	Exposure definition (higher figures represent better exposures)
0	Airflow fully blocked adjacent to sensor (within 1 m)
1	Airflow fully blocked at medium distance (1–4 m)
2	Airflow fully blocked at larger distance (4–10 m)
3	Airflow partially blocked adjacent to sensor (within 1 m)
4	Airflow partially blocked at medium distance (1–4 m)
5	Airflow partially blocked at larger distance (4–10 m)
6	Airflow clear with long fetch (>30 m) over ship
7	Airflow clear with fetch 10–30 m over ship
8	Airflow clear with fetch 1–10 m over ship
9	Airflow clear with short fetch (<1 m) over ship

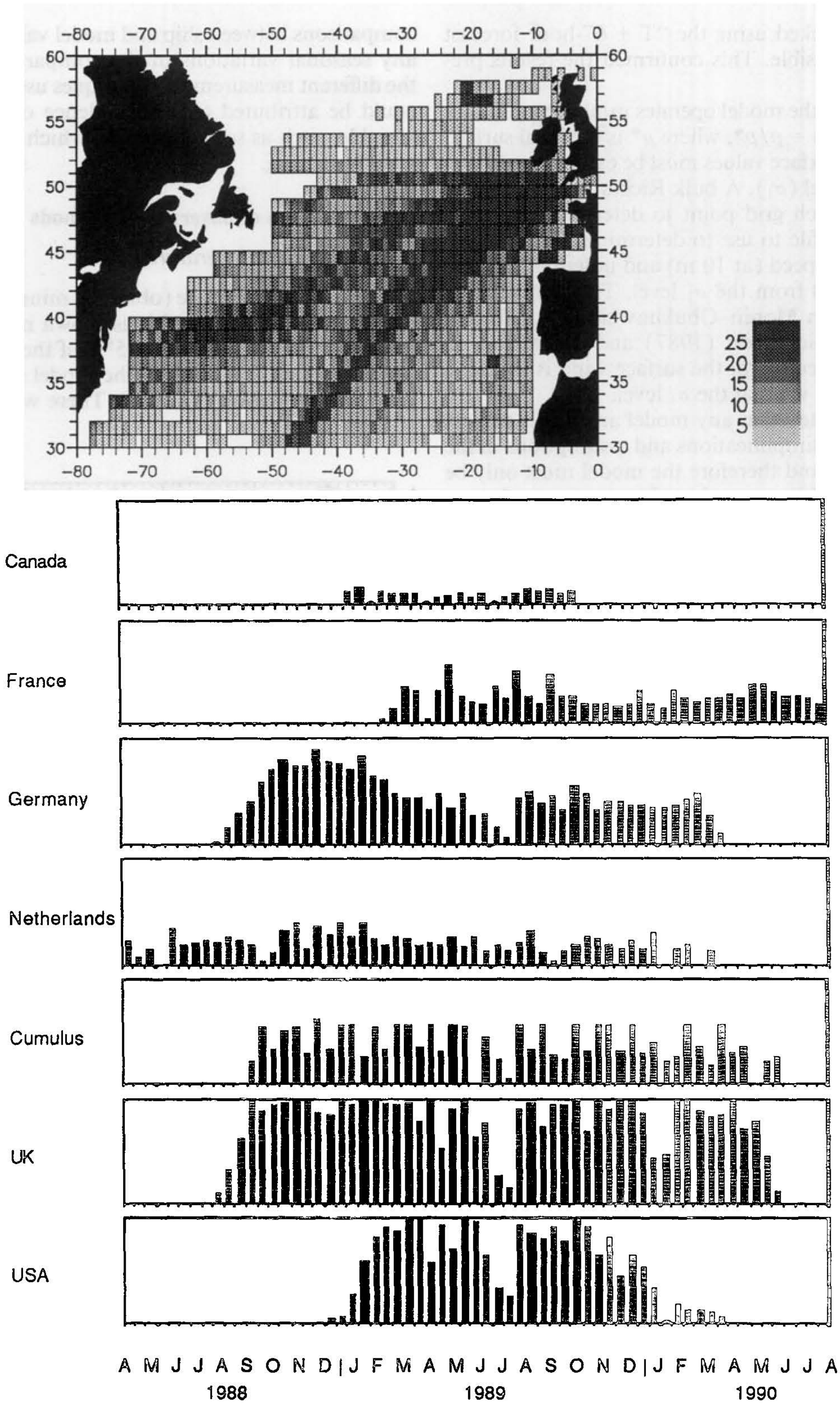

FIG. 2. (a) Map of the North Atlantic showing the distribution of the observations for the comparison dataset used for analysis. The scale shows the number of reports in each 1° latitude × 1° longitude area for the period of the project. (b) Number of VSOP-NA observations for the ships recruited by each of the participating countries. The vertical scale is 0–200 (or more) observations for each two-week period.

proach was checked using the "T + 6"-hour forecast fields where possible. This confirmed the results presented here.

Finally, since the model operates using σ coordinates in the vertical ($\sigma = p/p^*$, where p^* is the local surface pressure), the surface values must be extrapolated from the lowest σ level (σ_1). A bulk Richardson number is calculated at each grid point to determine the most appropriate profile to use to determine the "surface" values of wind speed (at 10 m) and potential temperature (at 1.5 m) from the σ_1 level. The profiles used are derived from Monin–Obukhov similarity theory; see Bell and Dickenson (1987) and Devrell et al. (1985). The direction of the surface wind is assumed to be that of the wind at the σ_1 level.

It must be noted that any model analysis is subject to errors due to simplifications and assumptions in the model physics, and therefore the model must only be regarded as providing a useful reference datum. In presenting the results we are not considering the absolute accuracy of the model's performance (although some conclusions can be drawn). Rather, using the model, we are comparing ship observations from widely differing areas and times to expose relative differences arising from the VSOP-NA ships' differing instrumentation and observing practices. Thus, for a variable V, the mean difference δV, was defined by

$$\delta V = \frac{\Sigma [V_O - V_M]}{N},$$

where V_O is the value observed by the VSOP-NA ship and V_M is the corresponding model value derived by linear interpolation from the four nearest model grid points. The summation is over all values of ($V_O - V_M$) corresponding to V_M values in some range V_M to ($V_M + \delta V_M$); N is the number of observations lying in that range. Where error bars are shown in presenting the results (section 3), these represent the standard error of the mean. These differences were analyzed by ship, by observing practice, and by prevailing meteorological conditions. One problem in this approach is that some VOS observations were used in the model data assimilation scheme; this will be considered later.

d. Geographical and temporal variations

The ships participating in the VSOP-NA project did not all provide observations throughout the observing period, and each ship reported from a restricted set of shipping lanes. Thus a requirement for the use of the model as a comparison standard is that any errors in the model analysis field should be spatially uniform and constant with time. As will be discussed, spatial variations were detected in the quality of the model fields, which hindered the comparisons for some variables. It was fortunate that, during the period of the VSOP-NA project, there were no major changes in the model coding. Seasonal variations were detected in the comparisons between ship and model values. However, any seasonal variations in the comparisons between the different measurement techniques used on the ships could be attributed to a dependence on some other variable, such as solar radiation, which itself was seasonally varying.

3. Comparison of observation methods

a. Sea surface temperature (SST)

The mean value of the (observed minus model) SST difference ΔT_S for each ship is shown in Fig. 3. Most values for ΔT_S lie within ± 0.5°C of the model value, which suggests that for SST the model analysis was a good representation of reality. There was a tendency

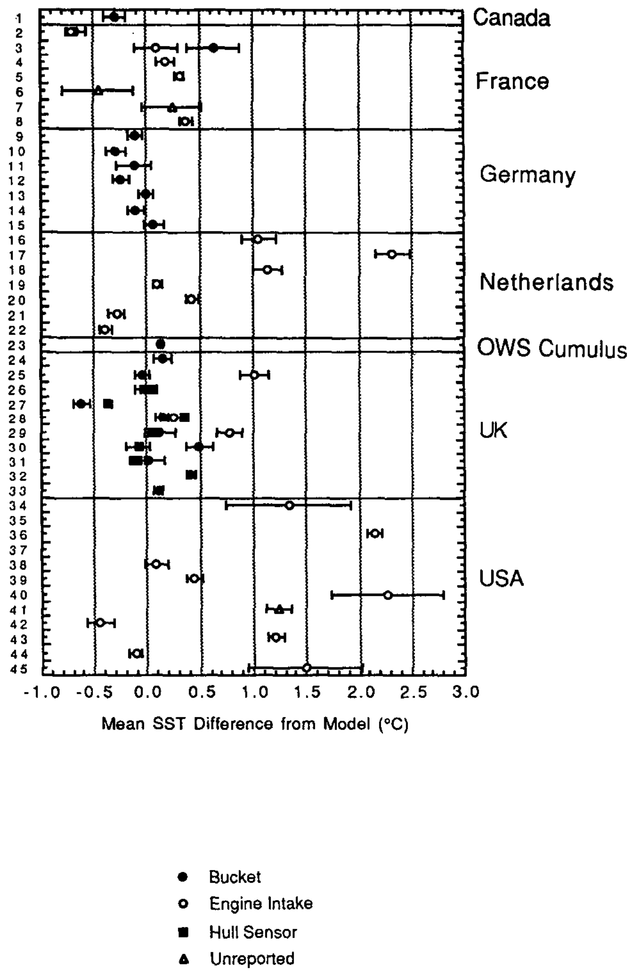

FIG. 3. Mean SST difference (°C) from model for each of the VSOP-NA ships (numbered 1–45 and grouped according to recruiting country). The symbols show the measurement method used (bucket, engine intake, hull sensor, or unreported); if more than one method was used on a ship the mean values for each method are shown separately. Ship 35 did not report SST and reports from ship 37 were on average 2.8°C less than the model SST (and therefore off the scale).

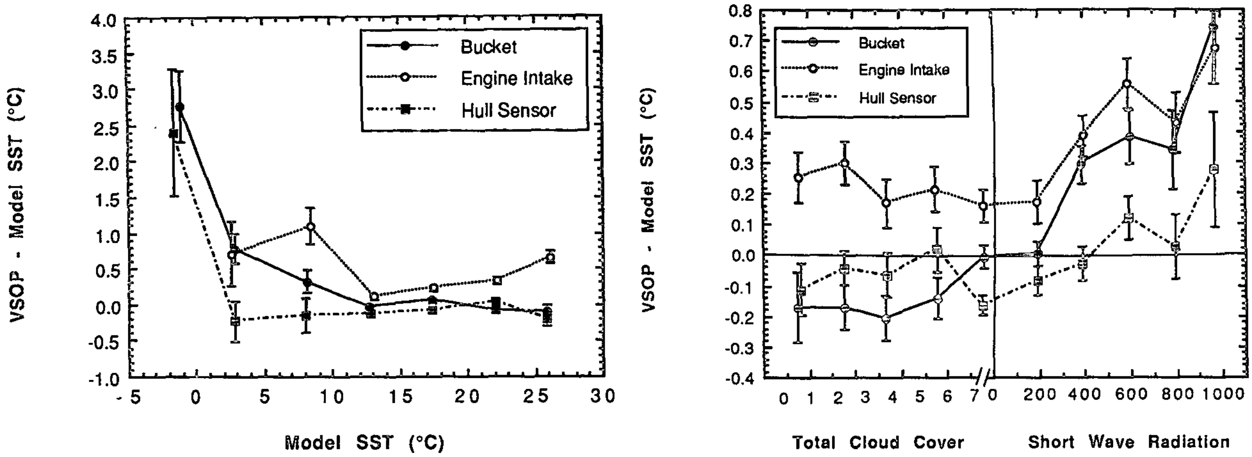

FIG. 4. (a) The mean difference in sea surface temperature measurements (°C, VSOP-NA ship-minus-model value) plotted against the model sea surface temperature value. The results for each measurement technique are shown separately. (b) As Fig. 4a but nighttime data are plotted on the left side of the graph against the observed cloud cover. Daytime data are plotted on the right of the graph against shortwave radiation (calculated from the cloud observations following Smith and Dobson 1985).

for the ships to report values warmer than the model; the mean difference was 0.10°C with a standard deviation of 1.62°C. Ships with the warmest mean ΔT_S values were those using engine intake measurements. In some cases this appears to have been a consistent bias since the scatter of observations was small. Differences evident in Fig. 3 between ships recruited by one country or another were mainly associated with the choice between the use of buckets or hull contact sensors (Germany and the United Kingdom) or engine intake values (France, the Netherlands, and the United States). Biases were particularly evident for some ships recruited by the Netherlands (although these values may have been affected by the short datasets available) and some ships recruited by the United States.

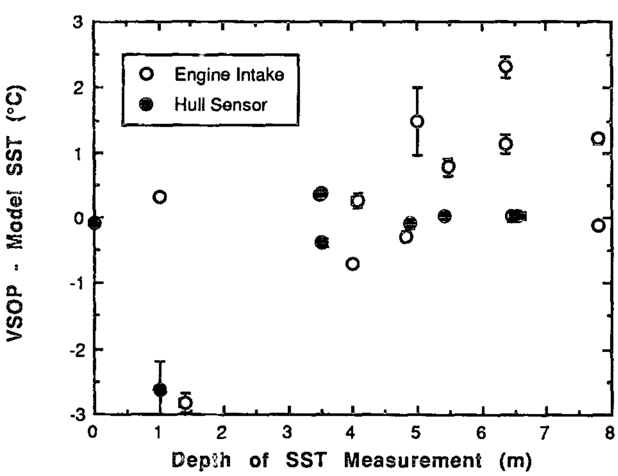

FIG. 5. The mean difference in sea surface temperature measurements (°C, VSOP-NA ship-minus-model value) for hull contact and engine intake methods plotted against the depth at which the SST was measured.

The mean variation of ΔT_S according to instrument type is shown in Fig. 4a plotted against model SST. Only a small number of observations from a few ships corresponded to model SST values below 5°C, and the large warm bias of the ship observations for model SST values below 0°C (that is, where the model would predict sea ice) cannot be considered significant. Few data existed at above 25°C. Between 5° and 25°C the hull sensor values were in close agreement with the model values, being about 0.2°C cold relative to the model at 5°C, and about equal near 25°C. The bucket values were also close to the model and hull contact values except in the lower SST range where they were relatively warm. The engine intake values were relatively warm over the entire range with an average bias of 0.3°C. Figure 4b shows the ΔT_S values plotted against cloud cover for nighttime values and solar radiation [calculated using the local sun time, the ships' observations of cloud cover, and the method of Smith and Dobson (1985)[1]] for daytime values. Whereas the engine intake values showed a relatively constant warm offset from the hull sensor values, the bucket values became significantly warmer with increasing solar radiation.

For hull contact and engine intake measurements, the variation of ΔT_S with measurement depth is shown in Fig. 5. With the exception of the single value at 1 m, the hull contact sensors showed no dependence on depth, whereas the engine intake values showed a possible trend toward warmer temperatures as the depth of measurement increased.

[1] A revised version of this model has been published (Dobson and Smith 1988), but use of this version would not be expected to significantly change the results published here.

b. Air temperature

The mean values of the (observed minus model) air temperature difference ΔT_a for each ship are shown in Fig. 6. Air temperature was measured on the VSOP-NA ships using a dry-bulb thermometer exposed in either a screen or a hand-held psychrometer. However, the ΔT_a values showed significant differences that appeared to depend more on the recruiting country than the type of instrumentation used; the reasons for this are not immediately clear but presumably depend on some aspect of national practice. Thus, psychrometer values from German recruited ships, and screen values from British ships were all up to 0.5°C lower than the model values. Psychrometer readings on the Netherlands recruited ships, and screens on the United States ships, were generally about 0.5°C warmer than the model. The French and United States ship air temperatures showed greater variation than those for other countries, with the French ships which used psychrometers giving higher air temperature values than other ships.

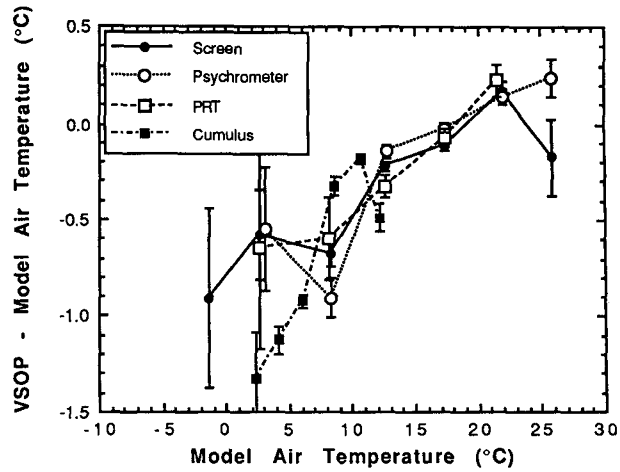

FIG. 7. The mean difference in air temperature measurements (°C, VSOP-NA ship-minus-model value) plotted against the model air temperature value. The results for each measurement technique are shown separately as are the data for the OWS *Cumulus.*

The mean values of ΔT_a are plotted as a function of the model temperature in Fig. 7. VSOP-NA air temperatures mainly corresponded to model values between 10° and 20°C. The French ships, which used the "Pommar" meteorological system with platinum resistance thermometers (PRT) in a small "stack of plates" screen, have been treated separately. The trend for all instruments was that the ship reports became warmer relative to the model as the temperature increased. This trend was also shown by the data from the OWS *Cumulus* suggesting that the trend was a feature of the model values. On average the psychrometer temperatures were warmest and the PRT coldest, but the differences were only of order ± 0.1°C.

There was also a geographical variation of ΔT_a. Figure 8 shows the average ΔT_a for observations during the winter night (to avoid solar radiation effects). The ship temperatures were colder relative to the model in a band running from about 55°N, 30°W to 30°N, 60°W. The amplitude of the variation was over 2°C. Most relatively warm ship values occurred to the southeast of this band. Because this geographical variation of ΔT_a cuts across many of the shipping routes (which determine the variations in sampling density), and also because the OWS *Cumulus* showed similar trends to the VSOP-NA ships, this geographical variation is believed to have been a characteristic of the model analysis rather than evidence of errors in the ship observations. Thus, although biases between the model and the ships became small when averaged over the whole area and period (the mean difference was −0.16°C with a standard deviation of 1.73°C) significant differences occurred over particular seasons or areas. That such variations in the model values may

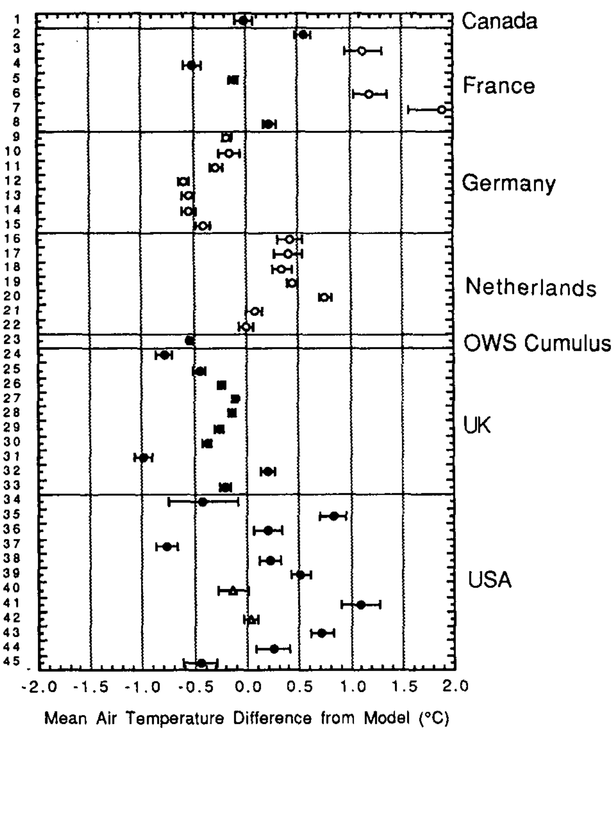

FIG. 6. Mean air temperature difference (°C) from model for each of the VSOP-NA ships (numbered 1–45 and grouped according to recruiting country). The symbols show the measurement method used (screen, psychrometer, or unreported).

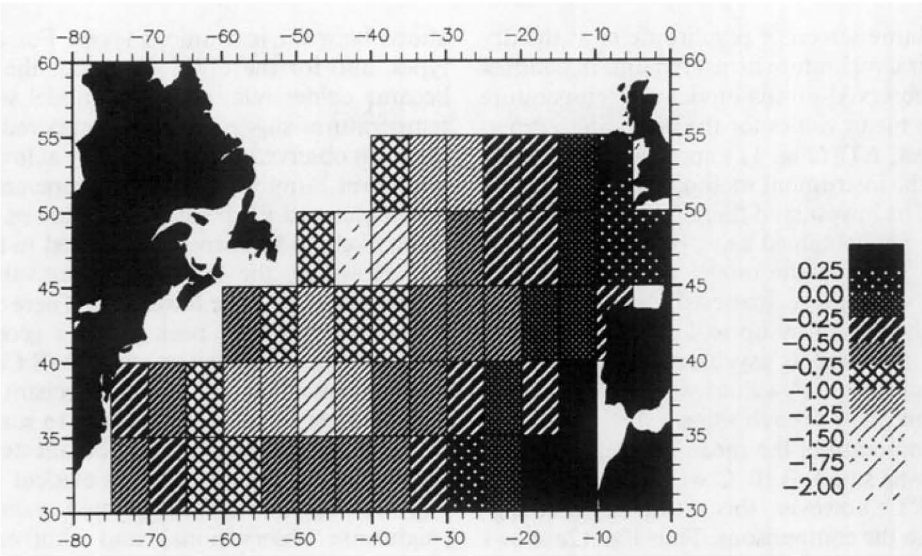

FIG. 8. Nighttime winter (November–March inclusive) air temperature difference (°C, VSOP-NA ship-minus-model value) averaged for 5° squares over the North Atlantic Ocean.

exist means that care is necessary in using the model as a comparison standard for air temperature.

Solar radiation affected both screens and psychrometers in a similar way. Figure 9 shows ΔT_a plotted against total cloud cover for observations at night and against shortwave radiation for observations during the day. Since at night the psychrometer values are about 0.2°C warmer, and since in the mean the psychrometers are warmer than the screens, it may be that the screens show a greater radiative effect by, at most, about 0.2°–0.3°C. There would also appear to be a model bias of about 0.5°C. During the day both types of sensors showed a similar trend with the mean ΔT_a rising to over 1.5°C when the radiation is high. The radiation error was not strongly dependent on the exposure of the instrument (Fig. 10). A large error occurred for a few very poorly exposed screen measurements, but overall the dataset was dominated by the data with medium and good exposure, which all showed a similar error. A dependence on the relative wind speed was found with stronger values of the relative wind decreasing the solar radiation effect.

c. Humidity

Humidity on the VSOP-NA ships is measured by wet-bulb temperature depression. The wet bulb is

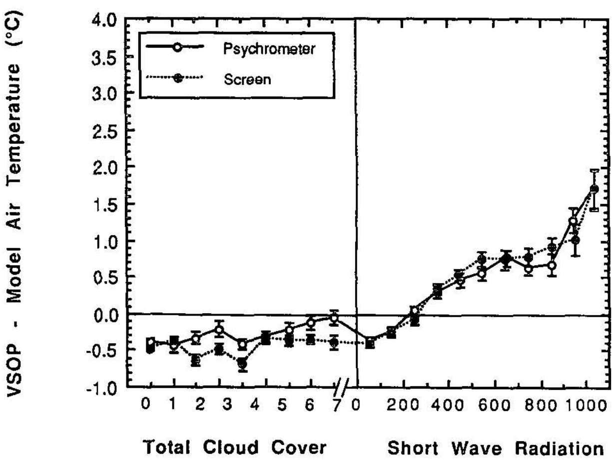

FIG. 9. The mean difference in air temperature measurements (°C, VSOP-NA ship-minus-model value) plotted against observed cloud cover for nighttime data (left side of graph) and against incoming shortwave radiation for daytime data (right side of graph). Data from screens and psychrometers are shown separately.

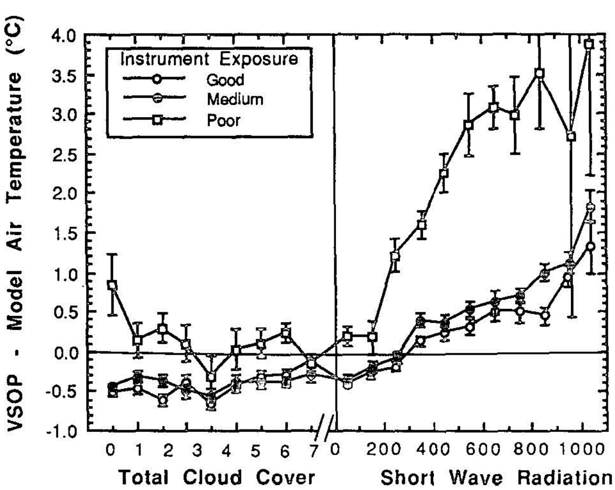

FIG. 10. As Fig. 9 but showing the effect of different instrument exposures. The exposure classes were "good" (6–9), "medium" (3–5), and "poor" (0–2), where the ratings are defined in Table 2.

housed in the same screen or psychrometer as the dry bulb used for air temperature measurement. In contrast to the mean observed-minus-model air temperature differences, the mean values for the dewpoint temperature differences, ΔT_d (Fig. 11) appeared to be more dependent on the instrument method than the country or ship route. The lowest, and therefore probably most reliable values, were obtained using psychrometers and were up to 0.5°C below the model values (Fig. 12a). Screen values were more scattered, being generally warmer than the model by up to 1.5°C. It is notable that, although the French psychrometer ΔT_a values were biased high, the ΔT_d values were similar to the screen values on other French ships.

As for air temperature, the mean difference for the entire dataset was small, 0.10°C with a standard deviation of 1.62°C; however, this mean disguised significant trends in the comparisons. Thus Fig. 12a shows that, plotted as a function of the model dewpoints, ΔT_d showed both a trend compared to the model and variations between instrument types. For all instrument types, and for the OWS *Cumulus,* the value of ΔT_d became colder relative to the model with increasing temperatures suggesting that compared to the model the ships observed higher humidity at low temperatures and lower humidity at higher temperatures. The psychrometer and PRT values were lower, and therefore more likely to be correct, compared to the screen values. However, the OWS *Cumulus* values, based on screen readings, were lower still. There is a possibility that this could have been partly a geographic effect; however, this comparison with OWS *Cumulus* values suggests that all the VSOP-NA merchant ship data may overestimate the humidity values to some extent.

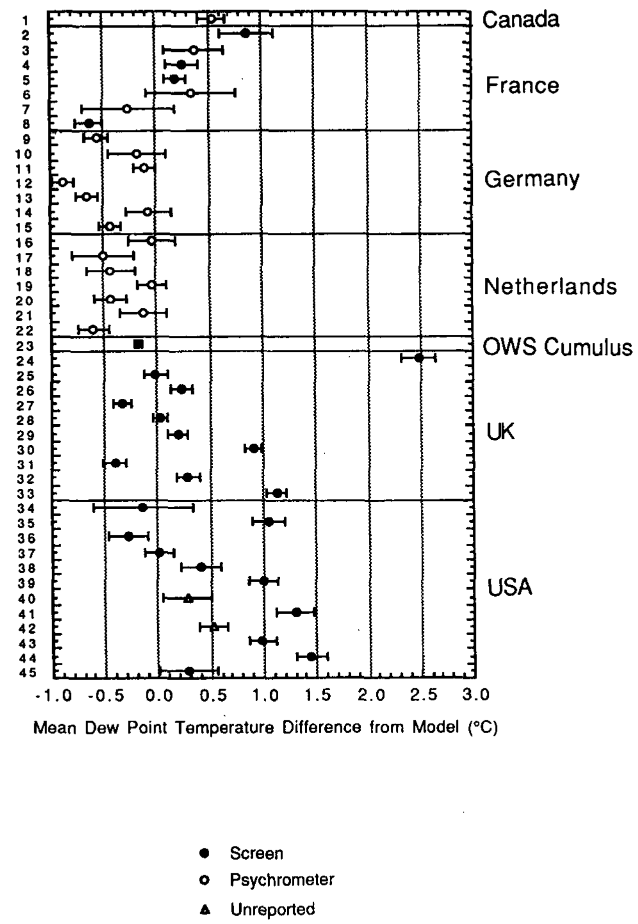

FIG. 11. Mean dewpoint temperature difference (°C) from model for each of the VSOP-NA ships (numbered 1–45 and grouped according to recruiting country). The symbols show the measurement method used (screen, psychrometer, or unreported).

Dewpoint data should not be affected by solar-radiation-induced errors. This is evident by comparing Fig. 12b, which shows ΔT_d plotted against cloud cover (nighttime observations) and shortwave radiation (daytime data), with the corresponding plot for air temperature (Fig. 9). Both screen and psychrometer ΔT_d values were higher for large cloud cover, and presumably higher humidity, conditions. These fluctuations were, however, relatively small compared to the radiation-induced error in air temperature.

d. Wind observations

Wind speed and direction reported by the VSOP-NA ships were obtained either by using fixed or hand-held anemometers measuring the wind relative to the ship, or derived from visual observations of the sea state and converted to wind speed using a Beaufort wind scale. The wind reports from the VSOP-NA ships were used by the model in calculating the analysis field. However, comparisons using the T + 6 forecast fields from the model were not significantly different from the results based on the model analysis results that are presented here.

For each ship the mean values of the observed-minus-model difference for wind speed Δv and direction Δd are shown in Figs. 13a and 13b. The Δv values were positive for all ships, regardless of instrument type; the mean difference was 2.9 kt[2] with a standard deviation of about 5 kt. The mean Δd values for most ships (Fig. 13b) were within ±5° of the model value. Six of the ten VSOP-NA ships that used wind vanes were among the seven ships with mean differences greater than ±5°. Even the OWS *Cumulus,* which uses wind vanes, showed a bias compared to the model of 5°. This probably illustrates the difficulty of aligning a wind vane with the ship's head to better than ±10°. Most individual Δd values, whether visual or instrument derived, were within ±10° of the model value. Outside this range some of these differences would have been due to small-

[2] Ships report winds in either knots or meters per second. To simplify comparisons with the ship speeds, this paper uses knots. To the accuracy required here, 1 kt = 0.5 m s^{-1}.

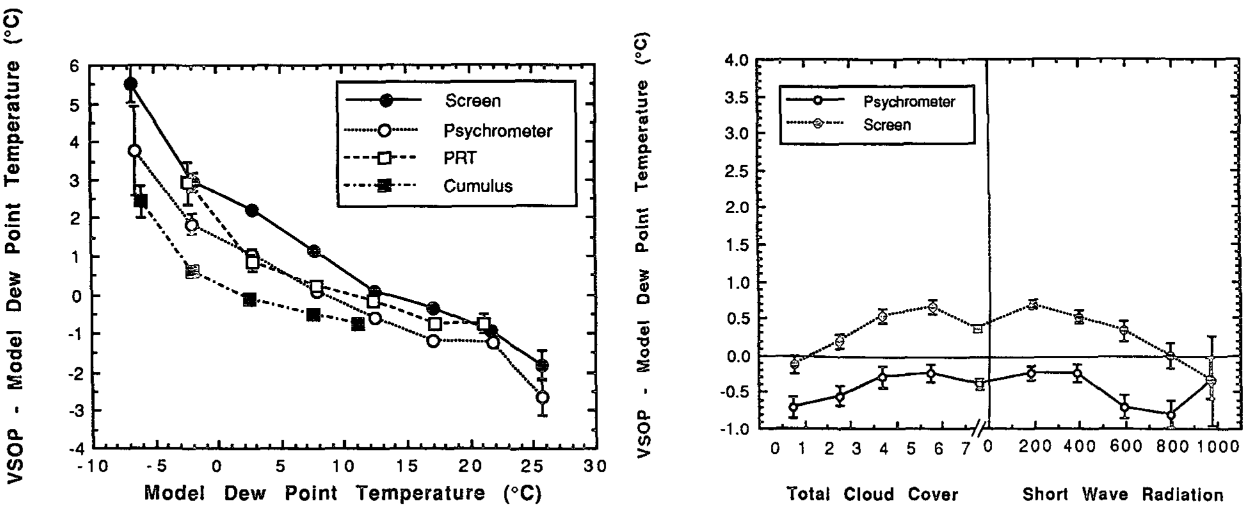

FIG. 12. (a) The mean difference in dewpoint temperature measurements (°C, VSOP-NA ship-minus-model value) plotted against the model dewpoint temperature value. The results for each measurement technique are shown separately as are the data for the OWS *Cumulus*. (b) The mean difference in dewpoint temperature measurements (°C, VSOP-NA ship-minus-model value) plotted against observed cloud cover for nighttime data (left side of graph) and against incoming shortwave radiation for daytime data (right side of graph). Data from screens and psychrometers are shown separately.

scale features of the wind field that are not represented in the model. However, the conversion of anemometer winds to true winds was also a significant source of error (see the following).

The variation of Δv, with model wind speed is shown in Fig. 14. The positive bias of the ship observations compared to the model is evident at all wind speeds. The distribution of model wind speed values corresponding to the VSOP-NA observations peaked in the range 10–15 kt. The most likely values were in the

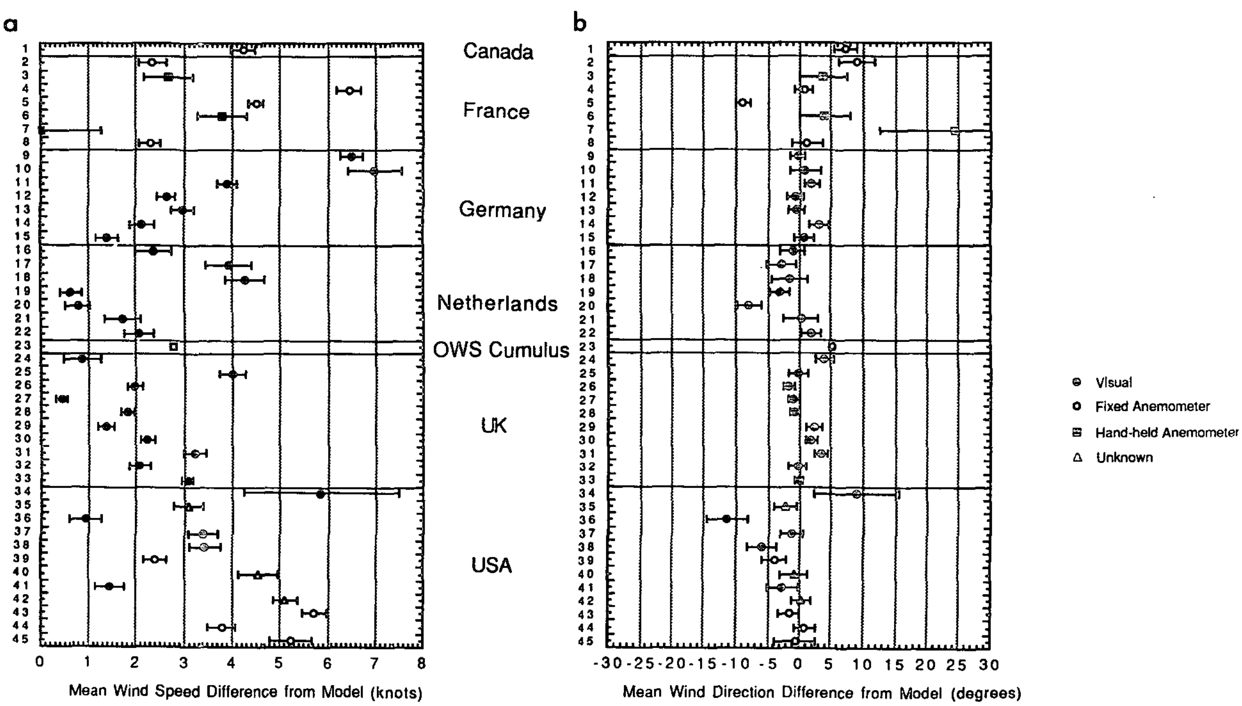

FIG. 13. (a) Mean wind speed difference (kt) from model for each of the VSOP-NA ships (numbered 1–45 and grouped according to recruiting country). The symbols show the measurement method used (visual, fixed anemometer, hand-held anemometer, or unreported). (b) As Fig. 13a but for wind direction (deg).

range 5–25 kt with almost all ships reporting wind speed values in that range. In this most commonly observed range, Δv increases with wind speed. There was a significant dependence on the estimation method used. Visual wind speeds were about 2.5 kt lower than winds from fixed anemometers for the most frequently observed wind speed range, but closer to fixed anemometer values at high wind speeds. Below 15 kt, winds from hand-held anemometers gave similar Δv values to visual winds, while at higher wind speeds the few observations that were obtained showed large scatter.

For those ships that used anemometers, errors might be expected due to anemometer height and in the calculation of true wind velocity. Following the recommendation of Dobson (1981), the anemometer values shown in Fig. 14 were not corrected for the height of the anemometer. For the VSOP-NA ships, variation of Δv with the anemometer height is shown in Fig. 15. This plot also shows the required correction to the wind speed to give 10-m values assuming neutral stability (Dobson 1981) and assuming a model offset of 2 kt. It would appear that the variation of the anemometer errors with height was possibly greater than would be predicted from the vertical wind gradient, possibly because of flow acceleration over the ships. If the VSOP-NA observations are corrected for anemometer height (Fig. 16), the Δv values are generally lower than the visual wind estimates. The anemometer observations from the OWS *Cumulus* were lower still; however, it is known that most wind observations are obtained with the *Cumulus* drifting downwind with the wind on the beam. Since no correction is applied for the ship's drift speed, it is possible that the OWS *Cumulus* underestimates the wind by 1–2 kt.

A further error arises because officers on ships using anemometers must perform the vector subtraction of the ships velocity from the measured relative wind. Given that the typical speed of the VSOP-NA ships was 16–18 kt, and that the most likely wind speed was in the range 5–20 kt, a large error can result if this calculation is not performed correctly. To determine the likelihood of this error, the VSOP-NA ships were requested to report ships' speed and head, and the relative wind speed and direction, in addition to the true wind values. Since most VSOP-NA wind data were visual observations, the number of these relative wind observations reported was small (about 2500). How-

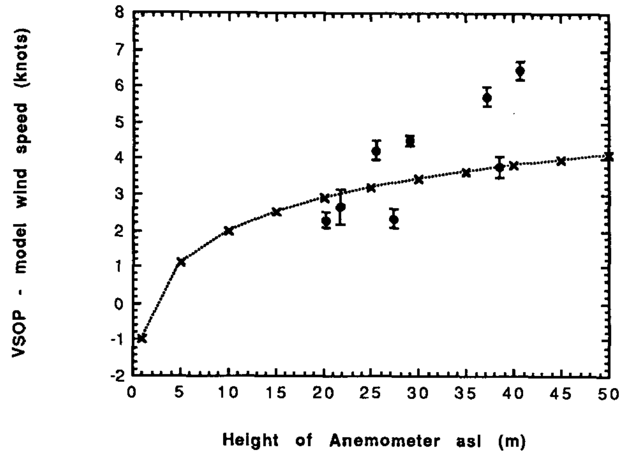

FIG. 15. The mean difference in wind speed measurements (kt, VSOP-NA ship-minus-model value) for each ship plotted against the anemometer height. The line indicates the correction from anemometer height to 10 m above sea level assuming a measured wind speed of 12 kt and assuming a 2-kt bias in the model values.

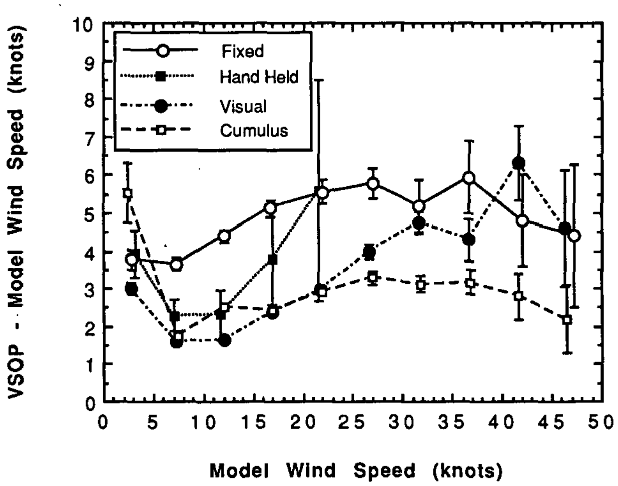

FIG. 14. The mean difference in wind speed measurements (kt, VSOP-NA ship-minus-model value) plotted against the model wind speed value. The results for each measurement technique are shown separately as are the data for the OWS *Cumulus*.

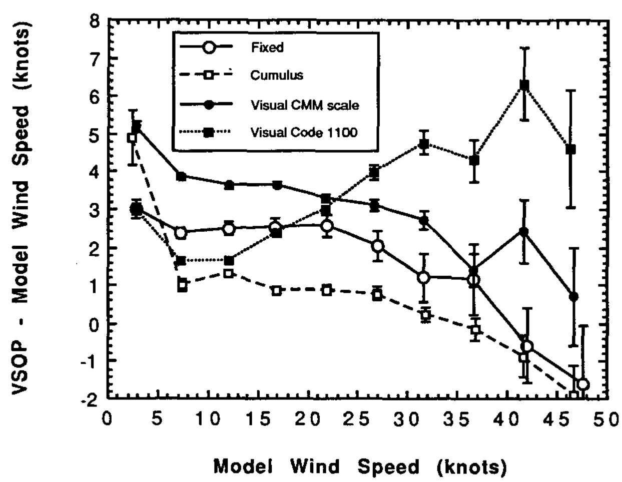

FIG. 16. The mean difference in wind speed measurements (kt, VSOP-NA ship-minus-model value) plotted against the model wind speed value. The results from fixed anemometers have been corrected for the anemometer height. The visual estimates have been corrected to the CMM Beaufort scale. (The dashed line represents the visual values using the code 1100 scale). Also shown are the anemometer data for the OWS *Cumulus*.

ever, using these data, the relative wind has been calculated from the reported true wind and compared to the observed relative wind. Only about 50% of the reported winds corresponded to calculated relative winds within ±2 kt of the observed value. A large fraction of the reports (30%) were more than ±5 kt different. For wind direction only 70% were within ±10°, and 13% were outside ±50°. These large errors in performing the vector calculation significantly and unnecessarily degrade the dataset of anemometer winds.

The accuracy of visual wind estimates depends on the Beaufort scale conversion used. That recommended for use on the VOS ships is known as "code 1100" but a different definition has been recommended by the World Meteorological Organization Commission for Marine Meteorology (CMM) as more accurate than code 1100 (WMO 1970). The advantages of the new scale were not considered sufficient to warrant the introduction of a new code on the VOS. Values for Δv calculated using the CMM scale are shown in Fig. 16. The visual winds are increased over the most often observed wind speed range but are reduced for winds over 20 kt.

When the visual data were divided into day and night values it was found (Kent et al. 1991) that winds above 15 kt were underestimated at night unless the ship also carried a fixed anemometer. Where a fixed anemometer was carried but visual winds reported, both day- and nighttime values showed similar characteristics to the daytime observations from ships that did not carry an anemometer. The atmospheric stability may affect the degree of white-capping and, hence, the appearance of the sea; however, no dependence of the visual winds on stability was found.

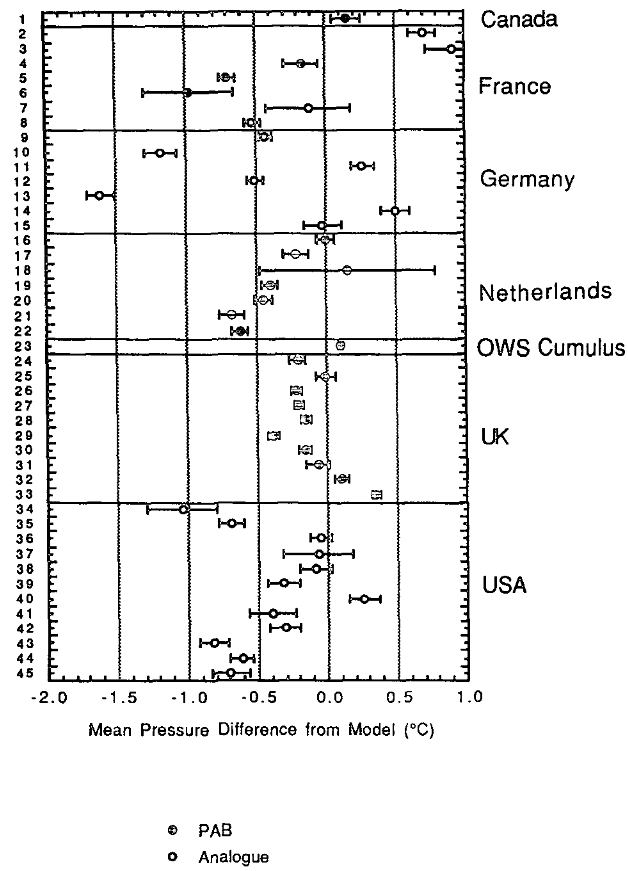

FIG. 17. Mean pressure difference (mb) from model for each of the VSOP-NA ships (numbered 1–45 and grouped according to recruiting country). The symbols show the measurement method used (PAB or analog).

e. Pressure

Pressure is measured on the VSOP-NA ships using either a precision aneroid barometer (PAB) or an analog aneroid barometer. None of the VSOP-NA ships used a mercury barometer. The mean values of the ship minus model pressure differences ΔP for each ship are shown in Fig. 17. For the majority of ships, ΔP was between 0.0 and −0.5 mb.[3] The Dutch and British ships used digital PABs and generally showed less scatter than ships recruited by the other countries that used analog aneroid barometers. Large (order 1 mb) but consistent biases occurred in the reports from some of the German and French ships.

Over the whole dataset the mean difference was −0.19 mb with a standard deviation of 1.96 mb. Examination of the scatter of ΔP values shows that for most ships there were some values that were probably incorrect by 10 mb. Such values would be easily removed by a quality control procedure that would reduce the standard deviation, probably to the 1.6 mb found by Hall et al. (1991).

Most reports corresponded to model pressure values between 1010 and 1030 mb with most ships contributing data in the range 990–1040 mb. For this well-sampled range of values the PAB measurements are closer to the model-derived pressures (Fig. 18). The differences are of order ±0.2 mb compared with ±0.5 mb for the analog pressures.

Ships' officers are required to correct the observation to sea level, taking into account variations in the height of measurement due to loading, and the air temperature (for air density effects). None of the VSOP-NA showed mean biases (of order 2–3 mb), which would indicate that no correction for the height of the barometer had been made. However, since the VSOP-NA ships reported the ship's draft with each observation it was possible to check whether variations in draft led to biases in the pressure observations. Figure 19 shows the mean observed-minus-model pressure difference (corrected by the mean bias for the ship providing each observation) plotted against the difference in the draft from that assumed in calculating the barometer height.

[3] 1 mb = 1 hPa.

If the ships had corrected for the varying draft then the data point should lie along the zero line. If no correction was made then the data points would lie along the sloping line. It would appear that, in general, no correction was made for changes in the reference height of less than 1 m, but that most ships corrected for variations in draft greater than 1 m.

The mean values of ΔP show a significant geographical variation (of order 0.3 mb) with the ships reporting higher pressures with respect to the model to the north, and lower pressures to the south. This geographical variation of ΔP cannot be easily explained by systematic errors that might be expected in the VSOP-NA observations, such as failure to correct for the air temperature, and therefore may be due to the model. Pressure observations are assimilated into the model and it would be expected that the mean difference between ship and model would have been small. Comparison with the "T + 6" forecast model suggested that the model had a tendency to underestimate the surface pressure, and that this was mostly, but not entirely, corrected by assimilation of the ship data.

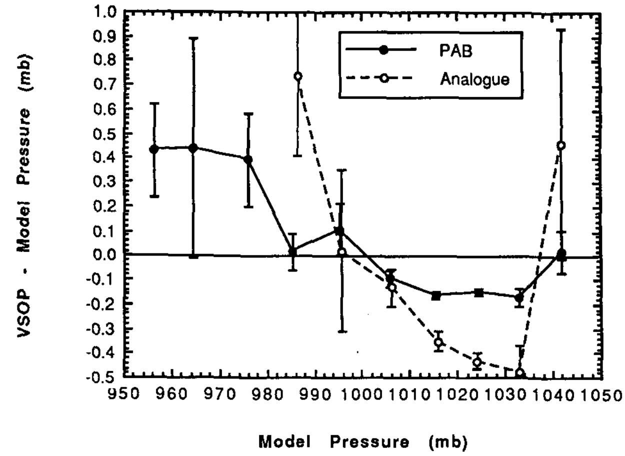

FIG. 18. The mean difference in pressure measurements (mb, VSOP-NA ship-minus-model value) plotted against the model pressure value. The results are shown separately for pressures measured by PAB and analog barometers.

4. Discussion

a. SST

The hull contact sensors, which are dedicated SST sensors attached to the inside of the hull plating, were only used on six of the ships recruited by the United Kingdom and were used for only 14% of observations. Analysis of these data showed, however, that compared to other measurement methods the hull sensors produced the most consistent SST values. The VSOP-NA Management Committee recommended[4] these sensors for use on the VOS and sample sensor designs were described in Kent and Taylor (1991).

Errors in bucket measurements may arise if the bucket is not immersed long enough to reach equilibrium, due to cooling while the bucket is hauled up to the deck, or if the water sample remains on deck for too long before the temperature is measured (Folland and Hsiung 1986). Indeed, bucket measurements are difficult or impossible in rough weather from large ships. The VSOP-NA values obtained using buckets were, however, generally comparable with the hull contact results. No clear relationship between the bucket ΔT_S values and the total turbulent heat flux was found, confirming the assumption of Bottomley et al. (1990) that the use of insulated buckets has reduced errors due to the bucket cooling during recovery or on deck to a negligible level. However, there was evidence that bucket readings were warm under conditions of high solar radiation. Since this effect did not vary with wind speed it was probably an error caused by the bucket heating on deck before use, rather than evidence for a near-surface warm layer in the ocean.

A warm bias of the engine intake values of on average 0.3°C relative to the bucket and hull contact sensors agrees with the findings of James and Fox (1972), who also found that this error increased with ship size. These authors assumed that the error was due to heat transfer into the cooling water between the intake and the point of measurement in the engine room and that this distance was greater on larger ships. Because many of the VSOP-NA ships were of similar length (about 200 m), it was difficult to obtain a clear relationship between the bias in the engine intake temperatures and the length of ship. However, the depth of the engine intake tended to be greater on the larger VSOP-NA ships, and the data did suggest, but by no means prove, that the temperatures measured from deeper engine intakes were relatively warmer. If real, this bias is likely to be due to the greater warming of the water temperature within the larger ships [as suggested by James and Fox (1972)] rather than stratification of the water column, since no dependence on wind speed was detected and the hull contact values showed no similar relationship with the sensor depth.

b. Air temperature

The analysis of the VSOP-NA data detected effects that were ascribed to varying biases in the model analysis. There was a geographical trend in the ship—model temperature differences of about 1°–2°C, and the model temperatures were biased low at low temperatures. To some extent, these variations hindered the use of the model as a reliable comparison standard.

[4] The results of the VSOP-NA were considered by the VSOP-NA Management Committee, which consisted of representatives of the different participating meteorological agencies; recommendations were made to the Commission for Marine Meteorology of WMO. These recommendations are listed in section 11 of Kent et al. (1991).

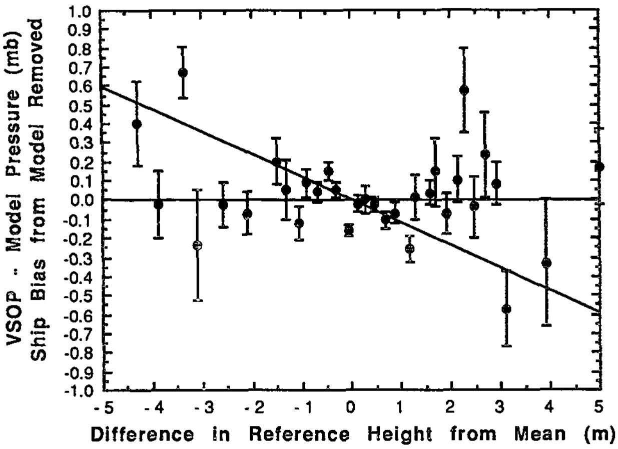

FIG. 19. The mean difference in pressure measurements (mb, VSOP-NA ship-minus-model value) plotted against the difference in draft at the time of observation from the ship's mean draft. The sloping line shows the expected relationship if no correction is made for variations in the ship's draft.

However, errors in the ship data were also clearly detected. The main measurement errors for the dry-bulb temperature are likely to be the "heat island" effect of the ship, the direct effect of solar radiation on the instrument, and allowing insufficient time for a psychrometer (normally kept in the wheelhouse) to reach outside temperature. These errors will increase the observed temperature value. However, when a dry bulb becomes wet through spray or rain, the ensuing evaporation will erroneously decrease the observed temperature reading. Thus biases in the ship-minus-model air temperature difference ΔT_a of either sign are possible and it is necessary to identify the cause of any bias detected.

The air temperatures from psychrometers on the German recruited ships and from the screens on the OWS *Cumulus* and the British ships all suggested that, in the mean, the model was biased warm compared to the ship observations by about 0.5°C. This bias was also detected when nighttime observations were examined (Fig. 9). Allowing for this bias, air temperatures from ships recruited by France, the Netherlands, and the United States were too warm by, in the mean, 1.0°–1.5°C. The VSOP-NA ship catalog (Kent and Taylor 1991) shows that the thermometer screens on some of the United States ships were very poorly sited, explaining the poor results for these ships. The warm bias for some of the French ships was probably an artifact due to these ships operating predominantly in the southeast of the region where the model was biased cold (Fig. 8). However, there is no obvious explanation for the psychrometer readings from the Netherlands ships being relatively high. Tests showed that it was not a geographical effect, and the sling psychrometers used on these ships were very similar in design to those used by the German ships.

On average, psychrometer measurements of air temperature were similar to, or slightly warmer (by about 0.1°C) than those from screens. The solar-radiation-induced biases for both types of instrument were similar, suggesting that the error is mainly caused by the heat-island effect of the ship rather than by direct solar heating of the instrument. A correction scheme for daytime marine air temperatures that depends on the solar radiation (deduced from the cloud cover) and the relative wind speed has been developed (Kent et al. 1993).

c. *Humidity*

For dewpoint observations, the main measurement errors result from inadequate ventilation or a contaminated or imperfectly wetted wick. Each will result in a decreased wet-bulb depression and, hence, increased dewpoint temperature. Thus the ship-minus-model dewpoint temperature difference values ΔT_d that indicate the ship observation being relatively high are more likely to be in error than those that show the ship's reading low. Thus, since the psychrometer humidity values were significantly lower by as much as 1°C compared to most of the screen values, the psychrometer observations were presumed to be more accurate.

It has already been noted that the OWS *Cumulus,* which uses screens, returned dewpoint temperatures even lower than the psychrometer readings, and that this suggests that all the VSOP-NA ships may have overestimated the atmospheric humidity to some extent. Since OWS *Cumulus* was away from the main shipping lanes, and since there were possibly geographical variations in the quality of model analysis fields, it is difficult to quantify this bias using the VSOP-NA results. However, assuming that the worst case is that the difference from the OWS *Cumulus* results in Fig. 12a represents an error in the VSOP-NA ship data, then this would be equivalent to a bias of about 7% in heat flux estimates from the psychrometer data and 20% for the screen data.

d. *Wind*

Considering first the anemometer derived winds, it was clear that hand-held anemometers did not give reliable results for model wind speeds above 15 kt. A significant source of error for all anemometers was the calculation of the true wind velocity from the relative wind. Since this is purely an arithmetic operation, this error in theory could be eliminated (and in practice much reduced) by providing the ships' officers with a correctly programmed calculator. A better solution might be to have the ships report relative wind velocity and ship velocity; however, this would require a significant change in the coding of ships' meteorological messages, which is to be avoided if possible.

Most wind observations corresponded to model winds below 20 kt. In this range the anemometer winds were on average about 2 kt greater than the reported visually estimated winds. However, correcting the anemometer winds for the height of measurement removed this bias so that the visual estimates were greater at most wind speeds. Adjusting the visual winds to the CMM Beaufort scale (WMO 1970) reduced this difference, particularly at the higher wind speeds, suggesting that CMM visual winds are more compatible with anemometer winds than estimates using the older code 1100 scale. This does not prove that the CMM scale is superior since errors may exist in the anemometer data because of ship motion, the blockage or acceleration of the airflow over the ship, and errors in visually averaging the dial readings. However, Cardone et al. (1990) have also demonstrated that adopting the CMM or similar scale does remove apparent climatic trends in the wind speed data caused by the increased use of anemometers. They too found that correcting anemometer-measured winds for an assumed anemometer height of 20 m improved the consistency of the climate dataset.

Even after making these corrections the ship wind speeds were in all cases higher than the model values. A contribution to this bias may come from miscoding or transmission errors, which can impose large positive, but not large negative, wind speed errors and may have had a significant effect, particularly at low wind speed. The OWS *Cumulus* Δv mean corrected to the 10-m height was about 2.6 kt, which was similar to the VSOP-NA mean difference. However, it has already been noted that the OWS *Cumulus* values may have been biased low by about 1–2 kt because of the failure to correct for the drift speed of the ship. Thus, overall, the data strongly suggest that the fine-mesh model may underestimate the wind speed by between 2 and 4 kt. Alves (1991) has discussed the surface momentum flux estimates from the fine-mesh model, and noted that the drag coefficient used is larger than recent empirical studies would suggest. Whether this causes the surface wind speed bias in the model detected in this study is not known.

e. Pressure

Pressure observations from ships using digital PABs appeared to be more accurate than those using analog instruments. A few instruments had a consistent calibration bias of order 1 mb. A major cause of mean differences between the ship observations of pressure and the model appeared to be a north-to-south variation in the model values of order 0.3 mb. Since this bias occurs over a distance of about 3000 km the implied error in the model-derived geostrophic winds would be negligible. In the mean the ship pressures were 0.2 mb lower than the model analysis. The lower pressures observed by the ships did not appear to be explained by a failure to correct for the height of, or changes in the height of, the barometer.

5. Summary

By using the analysis fields from an atmospheric forecast model as a comparison standard, the VSOP-NA project has identified various systematic errors in the meteorological observations from voluntary observing ships. These errors, mostly dependent on the method of estimation used, will cause biases in marine climate atlases and compilations of marine data such as COADS. Sea surface temperature data from engine intake thermometers were found to be biased high by on average 0.3°C. The dewpoint temperatures from fixed thermometer screens were biased high compared to psychrometer readings. The magnitude of this bias, of order 1°C, varied with the dewpoint temperature. There was, however, evidence that the psychrometer readings also may be biased by a similar amount. These humidity errors alone would cause an underestimate of the surface latent heat flux of up to 20%. Wind speeds from anemometers were biased high compared to visual winds by about 2 kt for winds up to about 25 kt. Use of the CMM version of the Beaufort scale rather than code 1100 reduced this difference significantly, but the VSOP-NA data do not indicate which is more accurate. However, compared to daytime values, visual winds at night were underestimated by about 1 m s^{-1} at 15 m s^{-1} and 5 m s^{-1} at 25 m s^{-1}.

The magnitude of the error in daytime air temperature measurements due to solar heating was about 1.5°C for an estimated incoming solar radiation of 1000 W m^{-2}. A similar bias affected both psychrometer and screen readings, and except for some very badly exposed screens, the bias did not vary greatly with instrument exposure. A correction scheme for daytime air temperature has been devised (Kent et al. 1993) based on the cloud observations and the relative wind at the time of observation.

The results of this study suggest that the most accurate dataset of VOS observations would be assembled by deriving corrections on a ship-by-ship basis by comparing observations with model-derived fields. It would be feasible to implement such a scheme within a weather forecasting center and thus quality control future VOS observations. For use with past observations, ship-by-ship correction may not be possible. Table 3 therefore summarizes the mean corrections that, based on the VSOP-NA results, should be applied to the VOS data from the North Atlantic. To the extent that ships operating in areas other than the North Atlantic will use similar observing methods and experience similar weather conditions to those studied here, then these corrections might be expected to hold for data from other areas. However, some of the errors depended on the instrument type and, hence, the country that recruited the VOS. Although many VOS operate world-

TABLE 3. Corrections to VOS observations based on the VSOP-NA results.

Variable	Correction
Sea surface temperature	Reduce engine intake temperature values by 0.3°C
Air temperature	Correct for solar radiation error (Kent et al. 1993)
Dewpoint	Linear correction to screen values 1°C at −5°C, 0°C at 26°C
Wind	Use CMM scale Increase nighttime visual winds

wide, it is likely that ships recruited by certain countries, and, hence, using particular observing techniques, will predominate in certain ocean areas. For this reason, correction of VOS data from other ocean areas on the basis of these North Atlantic results should be done with caution.

The need for corrections to the VOS data implies that estimates of the ocean to air fluxes of sensible and latent heat will also be in error. The corrections to the dewpoint, air temperature, and nighttime visual winds will tend to increase the heat transfer from sea to air. However the correction to engine intake temperatures and the adoption of the CMM wind scale will decrease the calculated heat transfer. Changes in the air–sea temperature difference will also change the fluxes due to the stability dependence of the transfer coefficients. Taking mean values for the whole area and period of the VSOP-NA project the changes in the implied fluxes were small. The mean sensible heat flux increased from 16 to 17 W m^{-2} and the mean latent heat flux from 94 to 97 W m^{-2}. However, the magnitude and sign of the implied flux changes varied greatly from area to area and season to season [for a more detailed discussion refer to Kent and Taylor (unpublished manuscript)].

The VSOP-NA project also identified certain types of instrumentation and observing methods that are preferable for future use on the VOS and recommendations were made by the VSOP-NA Management Committee to the CMM for implementation throughout the VOS system [see section 11 of Kent et al. (1991)]. For sea surface temperature, bucket measurements showed less scatter than engine-room intake values; however, bucket measurements are difficult to make from large ships, and hull sensors, which are being fitted to an increasing number of VOS, were recommended. For air temperature there was little to choose between psychrometers and screens (provided the latter are reasonably well exposed). However, psychrometers did provide lower, and presumably more accurate, dewpoint values. The VSOP-NA results do not show whether visual winds or anemometer winds are more accurate; however, if anemometers are used, they should be fixed rather than hand-held, and a reliable method of calculating the true wind velocity should be provided. The difference between anemometer and visual winds was minimized by correcting for anemometer height and using the CMM version of the Beaufort scale. Digital PABs gave more reliable results than analog instruments. Finally, the VSOP-NA analysis demonstrated that not only can an atmospheric forecast model be used as a comparison standard in assessing the quality of VOS data, it is possible to use the VOS observations to identify biases in the model analysis fields.

Acknowledgments. We must acknowledge the contribution to the VSOP-NA project of: the owners, masters, and officers of the VSOP-NA ships; the port meteorological officers, members of the VSOP-NA Management Committee, and staff of the Atmospheric Environment Service (Canada), Direction de la Météorologie Nationale (France), Deutscher Wetterdienst Seewetteramt, Royal Netherlands Meteorological Institute, Meteorological Office (United Kingdom), National Weather Service/NOAA (United States); and the World Weather Watch, World Climate Research Programme (WMO), and the Committee on Climate Changes and the Ocean (IOC/SCOR). The JRC analysis work was partially funded by the Ministry of Agriculture, Fisheries and Food.

REFERENCES

Alves, J. O. S., 1991: An assessment of the surface fluxes from the Meteorological Office numerical weather prediction models. Part 1: Momentum. *Meteor. Mag.*, **120**, 200–211.

——, 1992: An assessment of the surface fluxes from the Meteorological Office numerical weather prediction models. Part 2: Heat and fresh water. *Meteor. Mag.*, **121**, 59–84.

Bell, R. S., and A. Dickinson, 1987: The Meteorological Office operational numerical weather prediction scheme. Meteorological Office Scientific Paper 41, UK Meteorological Office, Bracknell, UK, 61.

Blanc, T. V., 1986: The effect of inaccuracies in weather-ship data on bulk derived estimates of flux, stability, and sea-surface roughness. *J. Atmos. Oceanic Technol.*, **3**, 12–26.

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: *Global Ocean Surface Temperature Atlas 'GOSTA.'* Meteorological Office/Massachusetts Institute of Technology, 20 pp. [+313 plates]

Bunker, A. F., 1976: Computations of surface energy flux and annual air–sea interaction cycles of the North Atlantic Ocean. *Mon. Wea. Rev.*, **104**, 1122–1140.

Cardone, V. J., J. G. Greenwood, and M. A. Cane, 1990: On trends in historical marine wind data. *J. Climate*, **3**, 113–127.

Devrell, C., F. Watkins, and J. Kitchen, 1985: *UKMO operational weather prediction scheme documentation paper no. 7.* UK Meteorological Office, Bracknell, UK, 15 pp.

Dobson, F. W., 1981: *Review of reference height for and averaging time of surface wind measurements at sea.* Marine Meteorology and Related Oceanographic Activities Report No. 3, World Meteorological Organization, Geneva, 64 pp.

——, and S. D. Smith, 1988: Bulk models of solar radiation at sea. *Quart. J. Roy. Meteor. Soc.*, **114**, 165–182.

Esbensen, S. K., and Y. Kushnir, 1981: *The heat budget of the global ocean: An atlas based on estimates from surface marine obser-*

vations. Tech. Rep. No. 29, Climate Research Institute, Oregon State University, Corvalis, Oregon, 27 pp.

Folland, C. K., and J. Hsiung, 1986: *Correction of seasonally varying biases in uninsulated bucket sea surface temperature data using a physical model.* Meteorological Office 13 Branch Memorandum 154, Meteorological Office, Bracknell, UK, 13 pp.

Hall, C. D., J. Ashcroft, and J. D. Wright, 1991: The use of output from a numerical model to monitor the quality of marine surface observations. *Meteor. Mag.,* **120,** 137–149.

Hellerman, S., and M. Rosenstein, 1983: Normal monthly wind stress over the world ocean with error estimates. *J. Phys. Oceanogr.,* **13,** 1093–1104.

Isemer, H.-J., and L. Hasse, 1985: *The Bunker Climate Atlas of the North Atlantic Ocean.* Vol. 1, *Observations.* Springer-Verlag, 218 pp.

——, and ——, 1987: *The Bunker Climate Atlas of the North Atlantic Ocean.* Vol. 2, *Air-sea Interactions.* Springer-Verlag, 252 pp.

——, and ——, 1991: The scientific Beaufort equivalent scale: Effects on wind statistics and climatological air–sea flux estimates in the North Atlantic Ocean. *J. Climatol.,* **4,** 819–836.

James, R. W., and P. T. Fox, 1972: Comparative sea surface temperature measurements. *Reports on Marine Science Affairs,* Report No. 5. World Meteorological Organization, 27 pp.

Jones, P. D., T. M. L. Wigley, C. K. Folland, D. E. Parker, J. K. Angell, S. Lebedeff, and J. E. Hansen, 1988: Evidence for global warming in the past decade. *Nature,* **332,** 790.

Kent, E. C., and P. K. Taylor, 1991: Ships observing marine climate: A catalogue of the Voluntary Observing Ships participating in the VSOP-NA. *Marine Meteorology and Related Oceanographic Activities,* Report No. 25. World Meteorological Organization, 123 pp.

——, B. S. Truscott, J. S. Hopkins, and P. K. Taylor, 1991: The accuracy of ship's meteorological observation—Results of the VSOP-NA. *Marine Meteorology and Related Oceanographic Activities,* Report No. 26. World Meteorological Organization, 86 pp.

——, Tiddy, R. J., and P. K. Taylor, 1993: Correction of marine air temperature observations for solar radiation effects. *J. Atmos. Oceanic Technol.,* **10,** in press.

Oberhuber, J. M., 1988: An atlas based on the 'COADS' data set: The budgets of heat, buoyancy and turbulent kinetic energy at the surface of the global ocean. Report 15, Max-Planck-Institute for Meteorology, 199 pp.

Smith, S. D., and F. W. Dobson, 1985: Estimation of solar radiation at sea. *The Ocean Surface: Wave Breaking, Turbulent Mixing and Radio Probing,* Y. Toba and H. Mitsuyasu, Eds., Reidel, 586 pp.

WMO, 1970: The Beaufort scale of wind force (technical and operational aspects). *Reports on Marine Science Affairs,* Report No. 3. World Meteorological Organization, 22 pp.

——, 1990: International list of selected, supplementary and auxiliary ships. WMO-47, World Meteorological Organization.

Woodruff, S. D., R. J. Slutz, R. L. Jenne, and P. M. Steurer, 1987: A comprehensive ocean–atmosphere data set. *Bull. Amer. Meteor. Soc.,* **68,** 1239–1250.

Wright, P., 1988: An atlas based on the 'COADS' data set: Fields of mean wind, cloudiness and humidity at the surface of the global ocean. Report 14, Max-Planck-Institute for Meteorology, Hamburg, 70 pp.