Assessing Air–Sea Surface Turbulent Heat Fluxes from Reanalyses Using Concurrent Shipboard Measurements

REBECCA FOODY,[a] SHENFU DONG,[b] MARLOS GOES,[b,c] RICARDO CAMPOS,[b,c] PEDRO PENA,[b] AND ZACHARY BARTON[b,c]

[a] *Cornell University Department of Earth and Atmospheric Sciences, Ithaca, New York*
[b] *NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida*
[c] *University of Miami Cooperative Institute for Marine and Atmospheric Studies, Miami, Florida*

(Manuscript received 25 February 2024, in final form 4 September 2024, accepted 23 October 2024)

ABSTRACT: Air–sea surface turbulent heat fluxes are crucial to forming accurate weather and climate predictions. However, there is considerable disagreement among current reanalysis products. The study assesses biases in these products using concurrent shipboard measurements of air temperature, barometric pressure, relative humidity, and upper-ocean temperature taken along two transbasin expendable bathythermograph (XBT) transects in the North Atlantic between 2020 and 2022. Following a quality control procedure, the data from shipboard measurements and hourly ERA5 and 6-hourly NCEP-2 reanalysis data were collocated to the locations and time of the XBT profile data. Our analyses show that mean biases among the three products are not statistically significant for all variables, but the biases for individual cases can be large. XBT could reliably estimate SST in the mixed layer, with a good comparison with the reanalysis products. Additionally, the correlation between the shipboard measurements and reanalyses is high ($r > 0.9$) for all variables except relative humidity (RH). Specifically, RH shows a correlation slightly lower between shipboard and ERA5 data ($r = 0.84$) and much lower between NCEP-2 and the other products ($r = 0.42$ for shipboard and $r = 0.46$ for ERA5 data). The collocated data were then used to compute the surface latent and sensible heat fluxes in four scenarios using the Coupled Ocean–Atmosphere Response Experiment flux algorithm version 3.6: in situ data with satellite winds, in situ data with ERA5 winds, ERA5 data, and NCEP-2 data. Overall, ERA5-based fluxes agree better with the shipboard estimates of both latent and sensible heat fluxes than NCEP-2-based fluxes, pointing to the importance of higher-resolution reanalysis to improve air–sea flux estimates.

KEYWORDS: North Atlantic Ocean; Heat budgets/fluxes; In situ atmospheric observations; In situ oceanic observations; Reanalysis data

1. Introduction

Turbulent fluxes have been widely applied to understand the air–sea interaction mechanisms, particularly in western boundary current regions (Frankignoul and Hasselmann 1977). The lead–lag covariance and correlation between turbulent heat flux (THF), sea surface temperature (SST), and SST tendency are used to identify the leading driver of SST fluctuations (Wu et al. 2006; Bishop et al. 2017; Bellucci et al. 2021). If the upper-ocean cooling is driven by the release of heat from the ocean into the atmosphere (atmospherically driven), SST and THF are expected to be in quadrature, and SST tendency and THF have a simultaneous negative relationship. Conversely, if the ocean advection (i.e., heat convergence from oceanic currents) is important for the SST change, such as in the Gulf Stream region, SST–THF zero-lag correlation is positive, and SST tendency is in quadrature with THF.

Therefore, to understand these relationships, which are necessary, for example, to reduce uncertainties in weather forecasts and seasonal-to-subseasonal predictions, there is a need for collocated upper-ocean and atmospheric observations. Current reanalysis data have improved considerably (Hoffmann et al. 2019), producing outputs on eddy resolving and hourly time scales, with their accuracy greatly enhanced due to model improvements and the evolution in the observing systems (Campos et al. 2020, 2022). Current assimilation systems ingest millions of surface observations per day (Hersbach et al. 2020), and remotely sensed data comprise a large part of them. Despite these improvements, there is still a significant spread of the surface heat flux estimates among data assimilation systems, and this spread is dominated by THF (Valdivieso et al. 2017). THF variability, and therefore biases, is particularly strong over western boundary current regions, where high-resolution data and reanalysis are essential for resolving air–sea interaction across the frontal regions such as the Gulf Stream (Kelly et al. 2010; Jin and Yu 2013). THF from commonly used reanalysis products is computed using bulk formulas driven by state variables such as winds, air temperature, relative humidity, and SST. Both the differences in the bulk formulas applied and the accuracy of state variables contribute to the spread of THF from different products.

Surface atmospheric observations from projects such as buoy arrays and the volunteer observing system (VOS), which collects along-transect high-resolution data from merchant ships around the world, are essential for the viability of these products and the calibration of satellite data (McPhaden et al. 2009). However, for scientific purposes, the VOS data have restrictions because they are only atmospheric—mostly sea level pressure and surface air temperature data—and are not

Corresponding author: Shenfu Dong, shenfu.dong@noaa.gov

DOI: 10.1175/JTECH-D-24-0028.1

© 2024 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

FIG. 1. Meteorological stations developed by NOAA/AOML and installed on the commercial ships (left) *Chicago Express* and (right) *Bremen Express*.

calibrated to climatic standards. Other projects, such as the Shipboard Automated Meteorological and Oceanographic System (SAMOS), use only refined data, which are more accurate and comprise more parameters, to be able to meet scientific standards (Smith et al. 2001, 2018).

As part of the Ships of Opportunity Program (SOOP) project, expendable bathythermographs (XBTs) are launched from ships following predetermined routes for decades (Goni et al. 2019). The XBT network measures the upper-ocean temperature (near surface to ~800 m) at high density (~25-km resolution), which monitors boundary currents, eddy structure, and meridional heat transport across basins. The potential for these data to be used for SST estimates and air–sea interaction has not been widely assessed, in part because of a lack of collocated surface atmospheric data.

In this work, we analyze surface atmospheric data collected since 2020 in the North Atlantic from meteorological stations, developed by the U.S. National Oceanic and Atmospheric Administration's (NOAA) Atlantic Oceanographic and Meteorological Laboratory (AOML) and deployed along two XBT transect ships. The potential to use these data for climate and air–sea interaction studies will be constrained by the quality of these data. Using these data, along with other observing systems such as satellite data, to estimate THF will allow an enhanced application of the XBT data, as well as a means to evaluate the quality of reanalysis data.

This paper is structured as follows. Section 2 describes the new meteorological data, the XBT transects along which these measurements were made, and the additional satellite and reanalysis data used. Section 3 describes the meteorological data corrections, the SST definition from XBTs, and the main formulas to calculate turbulent fluxes. Section 4 shows the results and discussion, and finally, section 5 concludes the paper.

2. Data

a. *Shipboard measurements*

The meteorological station (Fig. 1) was developed by the NOAA/AOML and measures surface air temperature (SAT; °C), relative humidity (RH; %), and barometric pressure (hPa). The sensors are manufactured by RM Young. According to the manufacturer, the model 4138VC temperature and relative humidity sensors provide an accuracy of ±0.3°C and ±2%, respectively, and the model 61302L barometric pressure sensor provides an accuracy of ±0.2 hPa. Wind speed (WS) is not measured in this meteorological station due to high turbulence on the surface of the ships that might interfere with accurate readings. Two meteorological stations are installed on the commercial ships *Chicago Express* and *Bremen Express*, which are part of the SOOP project that were recruited to conduct

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

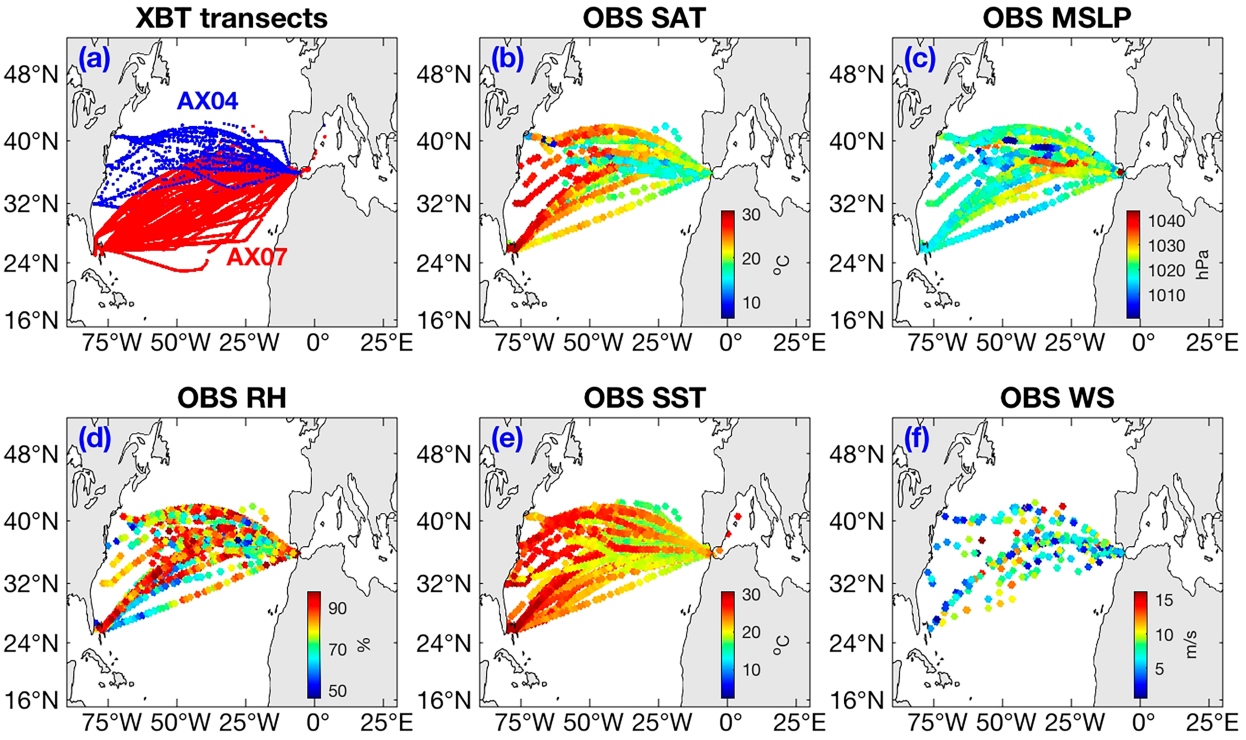

FIG. 2. (a) Location of ocean temperature measurements from XBT transects AX04 (blue, started in 2020) and AX07 (red, started in 1995). (b)–(f) Observed variables after quality control, collocation, and height corrections measured between 19 Aug 2020 and 23 Mar 2022. Variables include (b) SAT (°C), (c) MSLP (hPa), (d) RH (%), (e) SST (°C), and (f) WS (m s^{-1}).

the transbasin XBT transects AX07 and AX04 in the North Atlantic (Fig. 2a). XBTs are deployed at a high density of typically 25-km spacing between Miami and Gibraltar (AX07) and Gibraltar and New York (AX04). The meteorological stations were installed in 2020 and collected data constantly aboard the ships with 1-min temporal resolution. The stations were installed at about a 40–50-m height, and adjustments to standard height using key thermodynamic equations (e.g., hydrostatic balance and Clausius–Clapeyron equations) are required before comparing to other products (Stull 2017). The adjustments are based on the mean difference between station pressure measurements and reanalysis mean sea level pressure (MSLP) as described in section 3. To reduce the amount of data and increase the signal-to-noise ratio (Bourassa et al. 2003), the measurements are averaged to 30-min resolution after quality control and then collocated to the XBT deployment locations. The weather station data analyzed here were collected between 19 August 2020 and 23 March 2022.

In addition to the meteorological station data, SST data are inferred from XBT profiles. XBTs measure temperature profiles from near surface (0.67 m) to approximately 800 m, where depth is estimated using a fall rate equation. XBT probes take a finite time to adjust to the temperature of surrounding water after entering the ocean (Goes et al. 2017); therefore, it was suggested that near-surface temperature from XBTs (<3.7-m depth) should be treated with caution (Bailey et al. 1989; Kizu and Hanawa 2002). The temperature measurements at approximately 5-m depth are used as observed "bulk" SST in this study. The correction scheme for historical XBT data proposed by Cheng et al. (2014) is applied to all AOML XBT profiles, which reduces XBT temperature biases to ±0.04°C compared with conductivity–temperature–depth (CTD) sensors. The XBT data collected between August 2020 and March 2022 are used in this study.

b. *Satellite data*

To complement the shipboard measurements in the turbulent flux calculations, we use remotely sensed WS data from scatterometer (Ribal and Young 2020) and altimeter (Ribal and Young 2019) satellites from the Integrated Marine Observing System/Australian Ocean Data Network (IMOS/AODN). The resolution of altimeter data is approximately 1 Hz along track or 7 km (Ribal and Young 2019). The resolution of the scatterometer measurements within the broad ground-track swath is typically either 25 or 12.5 km (Ribal and Young 2020). The following scatterometer missions were selected: *ERS-2*, *MetOp-A*, *MetOp-B*, *Oceansat-2*, *QuikSCAT*, and *RapidScat*; the following altimeter missions were selected: TOPEX, *ERS-2*, *GFO*, *Janson-1*, *Envisat*, *Jason-2*, *CryoSat-2*, *HY-2*, *SARAL*, *Jason-3*, *Sentinel-3A*, and *Sentinal-3B*. The satellite along-track measurements were collocated with the XBT profile locations following the methodology of Campos et al. (2020, 2022) and Campos (2023), in which the data within a maximum space distance of 25 km and a time distance of 0.5 h were averaged to each location using a Gaussian weight function.

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

c. Reanalysis data

Reanalysis data are compared to the shipboard measurements. Two reanalyses are used: ERA5 from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Hersbach et al. 2020) and the National Centers for Environmental Prediction, version 2 (NCEP-2), reanalysis (Kanamitsu et al. 2002). The ERA5 is available hourly at 0.25° horizontal resolution, and the NCEP-2 is available at 2.5° horizontal resolution and 6-hourly intervals. The variables used are 2-m air temperature, MSLP, 2-m RH, 10-m WS, and SST. The data are collocated to the XBT profile locations according to section 3b.

3. Methods

a. Quality control

Quality control methods were utilized for the shipboard measurements. This included removing the following three categories of data that we deemed invalid.

1) GPS ERRORS

GPS errors include any data with locations not following a reasonable ship route that performed the AX04 or AX07 transects. This included both locations over land and some outliers throughout the transect. For the latter, using a typical ship speed of 20 kt (1 kt $\approx$ 0.51 m s^{-1}; 37 km h^{-1}), data that were less than an hour apart and separated by longitudes and/or latitudes greater than 1° or more were removed.

2) EXACT REPETITIONS

Exact repetitions of all variables for the weather stations, including location, time, and all atmospheric conditions, were removed.

3) UNPHYSICAL VALUES

All three observed variables from shipboard measurements were given data boundaries based on reasonable overall and seasonal values. Air temperatures of less than −10°C, along with December air temperatures greater than 30°C, were removed. Barometric pressures less than 870 hPa and greater than 1090 hPa were removed. Relative humidities less than 0% and greater than 100% were removed. Sudden spikes in values were likewise removed using a moving median process with a window of 21 points, where spikes exceeding three standard deviations from the mean of the window were eliminated. Our sensitivity test showed that many spikes were not successfully removed after applying the "despike" process; however, the spikes are mostly eliminated after repeating the despike process three times. Therefore, the despike process was applied three times for all atmospheric variables.

b. Collocation

The quality-controlled shipboard measurements and reanalysis products were then collocated in space and time to the XBT profiles' locations (i.e., weather station, satellite, and reanalysis data were only used in locations with XBT data). To do this, we (separately) averaged the shipboard and (ERA5 and NCEP-2) reanalysis data points that fell within the same hour and within ±0.5° latitude and longitude of the XBT data. Satellite data were already collocated to the XBT data as mentioned in section 2b. In total, there were 2870 XBT profile points; the number of valid weather station and satellite points was 1029 (MSLP), 1035 (SAT), 944 (RH), and 244 (WS), respectively. Together, quality control and collocation significantly reduced the amount of usable weather station data from over 5 million original data points.

c. Correlation between observations and reanalysis

The shipboard measurements were then compared to the ERA5 and NCEP-2 products. After the quality control and collocation, shipboard and ERA5 variables had correlations of 0.97, 0.93, and 0.76 for SAT, pressure, and RH, respectively. Shipboard and NCEP-2 variables, on the other hand, had correlations of 0.94, 0.91, and 0.37, respectively. Due to the high correlation between the shipboard measurements and ERA5, the observations with large differences from the ERA5 data were removed. This included observed SAT with differences larger than ±5°C, observed pressures with differences larger than ±8 hPa, and observed RH with differences larger than ±25%. Minimal data points were removed by those constraints, including 5 for SAT, 7 for pressure, and 18 for RH. Removing those large anomalies improved the correlations of shipboard measurements with ERA5 products to 0.98, 0.99, and 0.84 for SAT, pressure, and RH, respectively. The relationships between the shipboard measurements and NCEP-2 products are also improved, with correlations of 0.96, 0.98, and 0.42 for SAT, pressure, and RH, respectively.

d. Standardizing observations

Shipboard measurements were taken on the upper decks of the ships (Fig. 1) and therefore were not representative of surface level conditions. Therefore, the observed pressure was converted to MSLP using its relationship with reanalysis products, and the observed air temperature and relative humidity were adjusted to the same vertical level as the reanalysis products, i.e., 2-m height, using atmospheric thermodynamic principles. Below, we describe the adjustments applied to each of the measured (shipboard) variables. Formulas in this section are taken from Stull (2017).

1) BAROMETRIC PRESSURE

The correlation coefficient between the collocated observed pressure and the ERA5 MSLP was 0.93 and between the collocated observed pressure and the NCEP-2 MSLP was 0.91. Due to the high correlation and minimal outliers, the pressure was adjusted using its relationship with ERA5 MSLP. We calculated the mean difference between the shipboard and ERA5 pressure for each transect. This mean difference, ranging from 4.5 to 6.7 hPa per transect, was then added to each point of the observed pressure data along its respective transect. We should note that, because of the strong relationship among the three pressure products after removing large outliers, adjusting the observed pressure, SAT, and RH based on

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

mean pressure differences with NCEP-2 MSLP for each XBT cruise did not change the results presented in section 4.

2) AIR TEMPERATURE

To adjust the air temperature measurements to 2-m height, the mean pressure difference was first used to calculate the height of where the weather station was located on each transect. This height difference was found using the hydrostatic balance equation:

$$\frac{dP}{dz} = -g\rho, \tag{1}$$

where dP is the difference in pressure between two levels, dz is the height difference between two levels, g is the average surface gravitational acceleration (9.81 m s^{-2}), and ρ is the average air density at sea level (1.225 kg m^{-3}). Height differences were estimated to be between 37 and 56 m. The 2-m SATs were calculated using the following equation:

$$\text{SAT} = T_{\text{OBS}} + [\gamma \times (dz - 2\,\text{m})], \tag{2}$$

where T_{OBS} is the observed air temperature at each point and γ is the dry-adiabatic lapse rate (9.8 K km^{-1}). To account for the standard height occurring at 2 m above sea level, 2 m was subtracted from the calculated height differences dz. The dry-adiabatic lapse rate was used because no observed data points showed a relative humidity of 100%, and so we assumed that the air was not saturated in this range.

3) RELATIVE HUMIDITY

To calculate RH at 2 m, we first used the Clausius–Clapeyron equation to calculate the saturation vapor pressure at ship height e_s using the temperature at ship height T:

$$e_s = \exp[55.2966 - 6810.5245T - 5.0894\ln(T)], \tag{3}$$

Using the definition of RH, i.e.,

$$\text{RH} = \frac{e}{e_s} \times 100, \tag{4}$$

we solved Eq. (4) for the partial vapor pressure e at ship height. Then, e and the observed pressure P at ship height were used to calculate the mixing ratio r:

$$r = \frac{0.622e}{P - e}. \tag{5}$$

Here, r is assumed to stay constant between 2 m and the ship height. Rearranging Eq. (5), r is used to solve for e at 2 m utilizing P at 2 m [calculated in section 3d(1)]:

$$e = \frac{r \times P}{0.622 + r}. \tag{6}$$

We then used Eq. (3) to solve for e_s at 2 m, using the air temperature T calculated in section 3d(2). Finally, we found the RH at 2 m using Eq. (4), using e and e_s at 2 m.

e. *COARE 3.6 flux algorithm*

Direct observations of THF are scarce, highly localized, and infrequent. State-of-the-art bulk algorithms are commonly used to estimate these turbulent fluxes. However, previous studies (e.g., Blanc 1987; Eymard et al. 1999; Brunke et al. 2002; Moore and Renfrew 2002; Brodeau et al. 2017) have shown that discrepancies between bulk algorithms can lead to a relatively large spread in the computed turbulent fluxes. To avoid flux discrepancies stemming from different bulk algorithms, instead of comparing turbulent fluxes estimated from shipboard measurements with reanalysis flux products directly, in this study, we compute latent and sensible fluxes from the observed and reanalysis variables using the same bulk algorithm and perform analyses on those fluxes.

The latest version of the Tropical Ocean and Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE) flux algorithm COARE 3.6 (Fairall et al. 2003, 2011; Edson et al. 2013) is used to compute the latent and sensible heat fluxes. Latent heat flux (LHF) is calculated as follows:

$$\text{LHF} = \rho_a \times L_v \times C_E \times U_{10} \times (q_{\text{sea}} - q_{\text{air}}), \tag{7}$$

where ρ_a is the air density, L_v is the latent heat of vaporization, C_E is a turbulent exchange coefficient, U_{10} is the relative wind speed at 10 m (relative to ocean surface current), and $q_{\text{sea}} - q_{\text{air}}$ is the sea–air humidity difference (where the surface water specific humidity q_{sea} and the specific humidity of air q_{air} are derived using observed SST, P, and RH). The sensible heat flux (SHF) is calculated as follows:

$$\text{SHF} = \rho_a \times C_{pa} \times C_H \times U_{10} \times (T_{\text{sea}} - T_{\text{air}}), \tag{8}$$

where C_{pa} is the specific heat capacity of air, C_H is a turbulent exchange coefficient, and $T_{\text{sea}} - T_{\text{air}}$ is the sea–air temperature difference [where T_{sea} is the sea surface skin temperature derived from the observed bulk SST and T_{air} is equivalent to SAT calculated in section 3d(2)]. According to Eqs. (7) and (8), wind speed is significant to both LHF and SHF, RH is significant to LHF, and SST and SAT are significant to SHF. Turbulent fluxes computed from Eqs. (7) and (8) are directed into the atmosphere (ocean) when positive (negative).

To examine the individual contributions of the variables in Eqs. (7) and (8), and the effect that the in situ observations would have on the calculation of the turbulent fluxes in reanalysis products, a sensitivity test was performed. For this, the COARE 3.6 was run in four scenarios:

1) Observed variables (including weather station SAT, RH, and MSLP and XBT SST) with satellite winds
2) Observed variables with ERA5 winds
3) ERA5 variables
4) NCEP-2 variables

Therefore, scenario 1 is exclusively based on observations. Scenario 2 provides greater visibility of fluxes when considering the limited availability of satellite wind data for the locations and times of interest. ERA5 wind data were chosen because

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

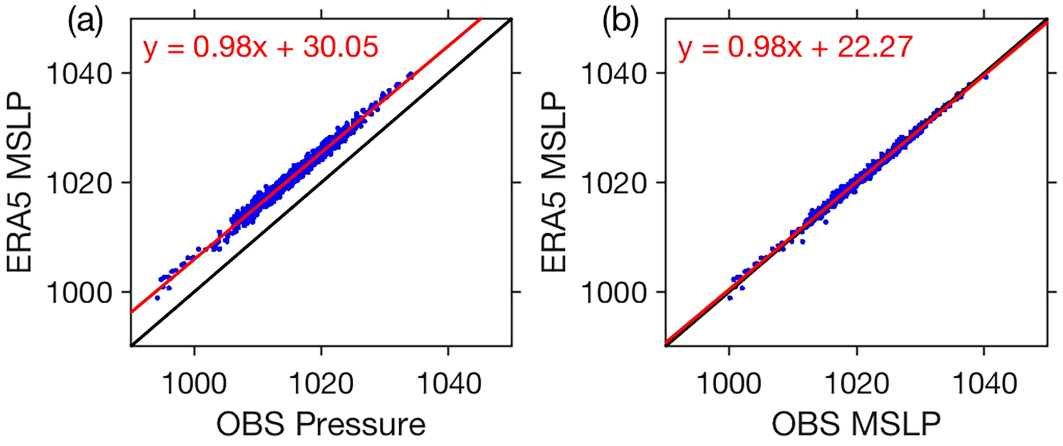

FIG. 3. Scatterplots of the observed pressure against ERA5 MSLP. (a) Observed pressure after quality control, collocation, and removing additional outliers but before height corrections. (b) As in (a), but the observed pressure has undergone height corrections and adjusted to SLP. The black lines represent the 1:1 line ($y = x$), and the red lines represent the line of best fit. Pressure is measured in hectopascals.

ERA5 and satellite winds have a correlation coefficient of 0.96, whereas NCEP-2 and satellite winds have a correlation of 0.85. Scenarios 3 and 4 use the same methodology as the previous two scenarios but utilize solely ERA5 and NCEP-2 reanalysis data.

We note that the 10-m wind speed is used as U_{10} in this study due to a lack of ocean surface current data at high spatial and temporal resolutions. To estimate the error associated with the missing surface currents, we calculated the THF using the climatological ocean surface current-derived surface drifters (Laurindo et al. 2017). On average, the ocean surface current speed in our study region is about 0.09 m s^{-1}, negligible compared to the averaged 10-m wind speed of 7 m s^{-1}. The THF with and without the climatological ocean surface current is very similar, with mean differences of 1.2 and 0.1 W m^{-2} for latent and sensible fluxes, respectively. The differences are mostly (~95% of the data) within ±5 W m^{-2} (5% of the flux value) for LHF and within ±1 W m^{-2} (5% of the flux value) for SHF. Because the ocean surface currents are set to zero in all four scenarios, the missing ocean surface currents do not impact the flux comparisons and conclusions presented in sections 4 and 5.

4. Results

a. Comparison of variables

Here, we investigate in more detail the agreement between reanalysis, satellite, and in situ variables. The variables analyzed are SLP, SAT, RH, SST, and WS, which are all corrected, quality controlled, and collocated with the XBT data as stated in section 3. As stated in section 3d(1), after removing outliers, but before performing any height corrections, the correlation coefficients of the collocated observed pressure with ERA5 and NCEP-2 MSLP were very high, with values of 0.99 and 0.98, respectively. After performing the height correction, the correlation coefficient between the observed and ERA5 MSLP was close to perfect, rounded to 1.00 (0.997). The correlation coefficient between the observed and NCEP-2 MSLP remained the same at 0.98. Figure 3 demonstrates the effect of this height correction; particularly, Fig. 3a (no correction) shows a consistent gap between the nearly parallel lines of best fit (with a slope of 0.98) and 1:1 line (i.e., zero bias line) of approximately 5–6 hPa, whereas Fig. 3b (with correction) shows minimal, if any, gapping.

While ultimately very similar in MSLP, the differences between the ERA5 and NCEP-2 products were more pronounced in the SAT and RH variables, which is at least partially driven by the difference in spatial and temporal resolutions. The time and space resolution of ERA5 (NCEP-2) data were hourly (6 hourly) and 0.25° (2.5°), respectively. The ERA5 product is therefore better equipped to represent the full range of values throughout the day and along the ships' paths. The height adjustment induced SAT changes between 0.34° and 0.53°C, with a mean increase of 0.43°C. The RH reduced by 1.7% on average due to the height adjustment, with values ranging from −2.5% to −0.8%. The height adjustments did not change the relationships of the observed variables with the reanalysis products. The correlation coefficients of the collocated observed air temperature with ERA5 and NCEP-2 SAT were 0.98 and 0.96, respectively, and both values held the same as before height corrections. Similarly, the correlation coefficients of the observed RH with ERA5 RH and NCEP-2 RH remained at 0.84 and 0.42, respectively.

To visualize the spatial scale of this analysis, Figs. 2b–f show color maps of the five in situ variables after quality control, collocation, and height corrections. The SAT and SST maps show a tendency for warmer temperatures in the western Atlantic Ocean, where warm ocean currents are directed from the tropics to subpolar regions. The mean and median observed MSLP is 1021 hPa. The observed RH appears to lack any notable spatial pattern. The amount of data available for satellite WS collocated with XBT profiles is strongly limited within the spatial and temporal boundaries used in this study.

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

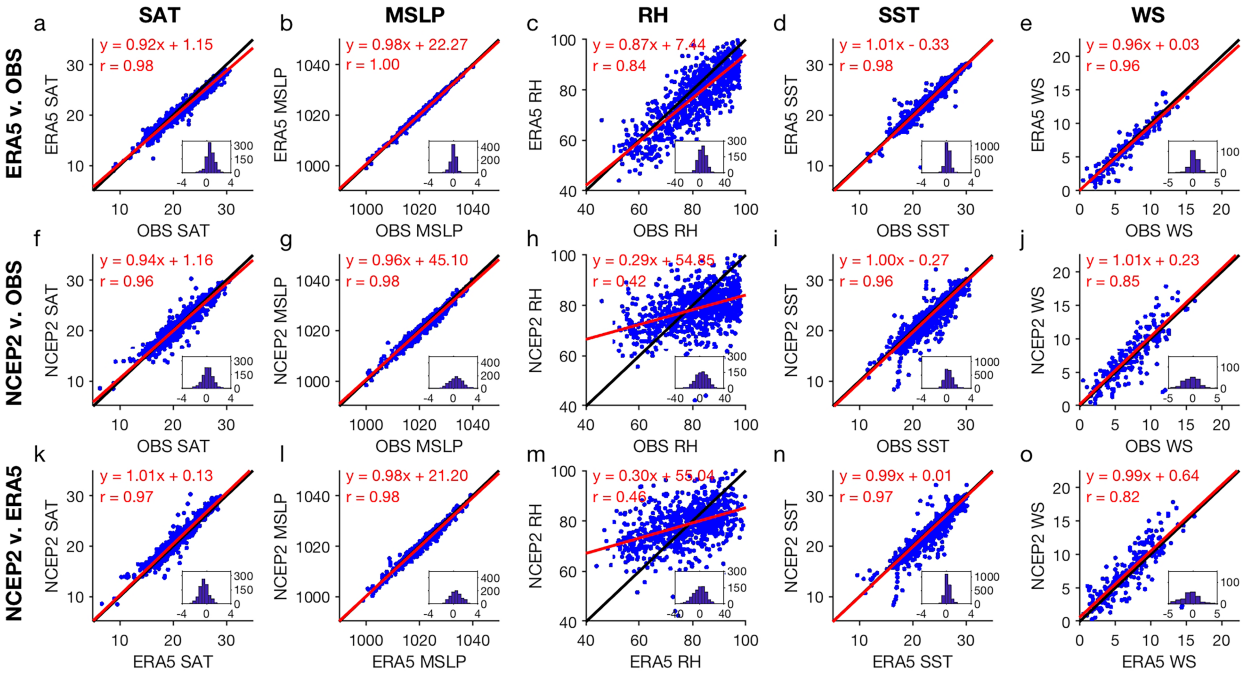

FIG. 4. Scatterplots comparing the observed SAT, MSLP, RH, SST, and WS with reanalysis products. Comparison of (first row) observed and ERA5 variables, (second row) observed and NCEP-2 variables, and (third row) reanalysis variables. The red lines represent the best linear fit lines, and the black lines represent the 1:1 line. The variables and their units are as follows: SAT and SST (°C), MSLP (hPa), RH (%), and WS (m s^{-1}). For visibility and continuity, some graphs exclude large outliers. Histograms represent the distribution of differences between products (first row) in the order of OBS − ERA5; (second row) in the order of OBS − NCEP-2; and (third row) in the order of ERA5 − NCEP-2. The *x* axes of the histograms represent the absolute difference in the two variables (in the units used in their associated scatterplots), and the *y* axes of the histograms represent the number of occurrences of these differences between the given variables. Regression relationships and the correlations *r* are given in each plot for the corresponding variables.

Next, we compare the observed SAT, MSLP, RH, SST, and WS with variables from the two reanalysis products, ERA5 and NCEP-2 (Fig. 4). Figure 4 demonstrates stronger relationships between observations and ERA5 variables as compared to the relationships between observations and NCEP-2 data. For all products and variables, except RH, the best fit slopes are above 0.9, in fact close to 1.0. For RH, despite a relatively strong correlation between observations and ERA5 (0.84) and a corresponding best fit slope of 0.87, there is still a significant spread along the 1:1 line (Fig. 4c). The histogram of RH differences (OBS−ERA5; inset plot in Fig. 4c) shows that ERA5 tends to underestimate the observed RH, with 69% of the available data showing negative biases. The regression line indicates that the negative biases in ERA5 RH tend to increase with increasing observed RH. However, the mean bias of −3.5% in ERA5 RH is not statistically significant considering the standard deviation of 6.4% in the RH differences. The correlation between the observed and NCEP-2 RH decreases considerably (to 0.42); this likewise occurs with the best fit slope, which decreases to 0.29 (Fig. 4h). Although the mean bias in NCEP-2 RH is relatively lower, about −2.9%, the differences with observed RH are more spread out with standard deviation of 10.5% (histogram in Fig. 4h). NCEP-2 tends to overestimate RH when the observed RH is lower than 80% and underestimate RH when the observed RH exceeds 80%. About 60% of the NCEP-2 RH data show negative biases. We also see a high discrepancy between the ERA5 and NCEP-2 RH products, which have a correlation of 0.46 and a best fit slope of 0.30 (Fig. 4m). The histogram in this panel shows that the RH differences between ERA5 and NCEP-2 products are more evenly distributed between positive and negative values.

Furthermore, although there tends to be strong relationships between all SST products, Figs. 4i and 4n highlight an interesting outlier in the NCEP-2 product that strongly underestimates the SST by almost 10°C. Those large biases in NCEP-2 SST occur near the eastern boundary as will be shown later. For both ERA5 and NCEP-2 SST, two-thirds of the data show cold biases compared to the observed SST and the other one-third shows warm biases. On average, the ERA5 SST has an insignificant cold bias of −0.17° ± 0.67°C. The cold bias in NCEP-2 SST is somewhat larger, −0.43° ± 1.10°C, but still insignificant. Figures 4d and 4i indicate a larger spread of the biases in NCEP-2 SST compared to the biases in ERA5 SST, consistent with the fact that only 4% of ERA5 SST data have absolute biases exceeding 1.0°C, whereas there are 21% of NCEP-2 SST data with absolute biases exceeding 1.0°C.

Reanalysis products also tend to show a small cold bias in SAT (Figs. 4a,f). Different from SST, the mean cold bias in ERA5 SAT (−0.64° ± 0.74°C) is larger than the mean bias in NCEP-2 SAT (−0.20° ± 1.10°C). However, similar to the SST, the differences between the observed and NCEP-2 SAT

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

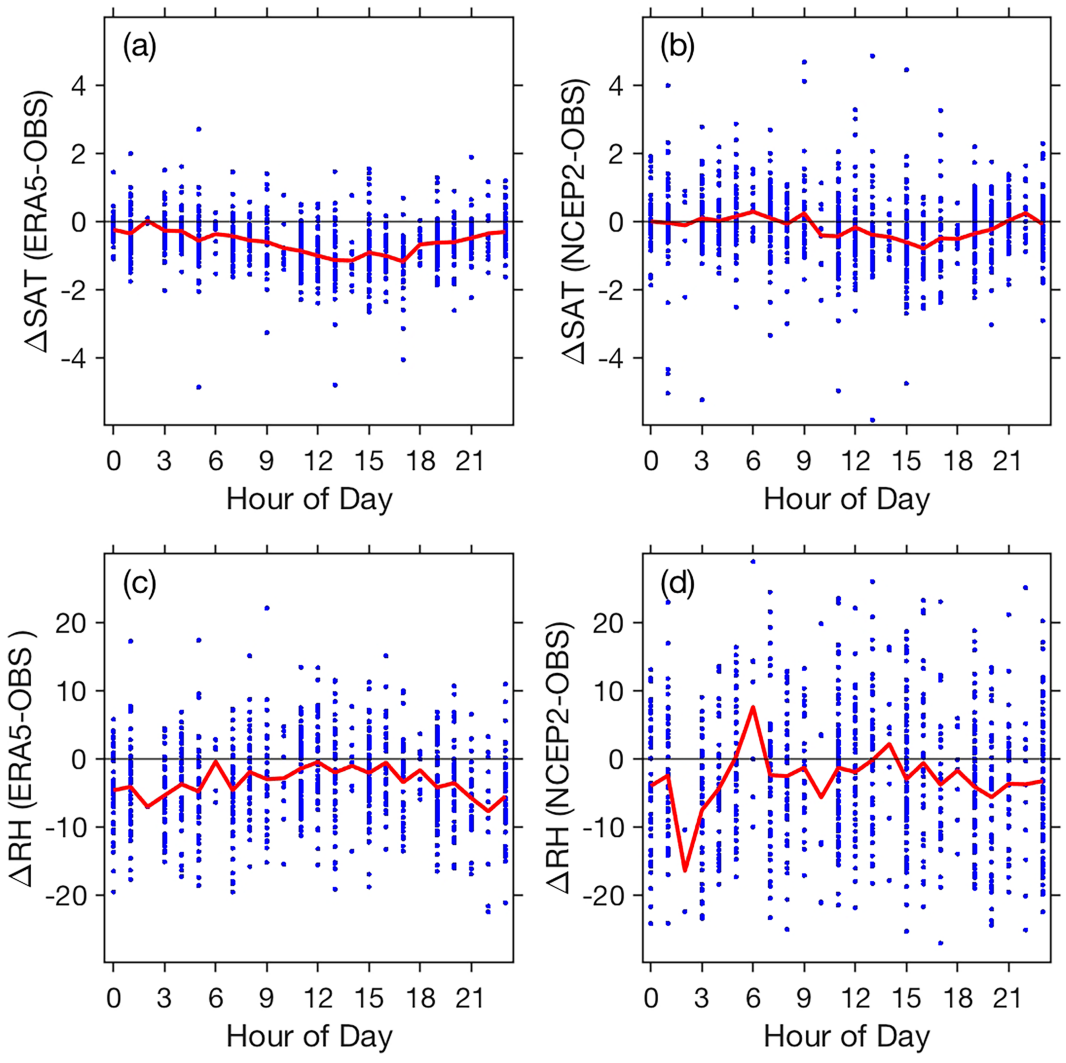

FIG. 5. (a),(b) Differences in SAT (°C) between observations and reanalysis products against the hour of the day: ERA5 − OBS in (a) and NCEP-2 − OBS in (b). (c),(d) As in (a) and (b), but for the differences in RH (%): ERA5 − OBS in (c) and NCEP-2 − OBS in (d). The red lines represent the mean difference for each hour.

show a slightly larger spread as compared to the differences between the observed and ERA5 SAT (inset histogram plots in Figs. 4a,f). It is also interesting that ERA5 SAT is dominated by cold biases, with 86% of the data showing cold bias and 14% showing warm bias. NCEP-2 SAT also has more cases with negative biases, but not as extreme as ERA5, with 61% of the data showing cold biases and 39% showing warm biases.

Similar to SST and SAT, all three WS products show strong relationships with each other (Figs. 4e,j,o). The biases in ERA5 WS are small, mostly within 1 m s^{-1} (Fig. 4e), whereas the biases in NCEP-2 WS are more spread out (Fig. 4j). The mean biases in both reanalysis WS are minimal with negative bias in ERA5 WS (-0.25 ± 0.97 m s^{-1}) and positive bias in NCEP-2 WS (0.21 ± 1.95 m s^{-1}). Comparisons of all three MSLP products (Figs. 4b,g,i) give the same conclusion, a slightly higher correlation between the observed and ERA5 MSLP than the correlation between the observed and NCEP-2 MSLP, and the differences between the observed and NCEP-2 MSLP have a larger spread than the differences between the observed and ERA5 MSLP.

We further examined the space and time dependence of biases in the reanalysis products. One possible explanation for the cold bias in SAT across products stems from the differences in SAT over the course of the day (Figs. 5a,b). In both panels, the bias is weakest overnight and strongest during midafternoon. As mentioned earlier, ERA5 SAT presents a dominant cold bias with time-of-day averaged differences (ERA5 − OBS) ranging from 0.01° to −1.19°C. The time-of-day averaged differences between NCEP-2 and OBS SAT (NCEP-2 − OBS) vary between 0.28° and −0.80°C. Both cold biases (particularly those in the afternoon) are speculated to be due in part to the resolution of the reanalysis products that prevent true maxima and minima from being able to

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

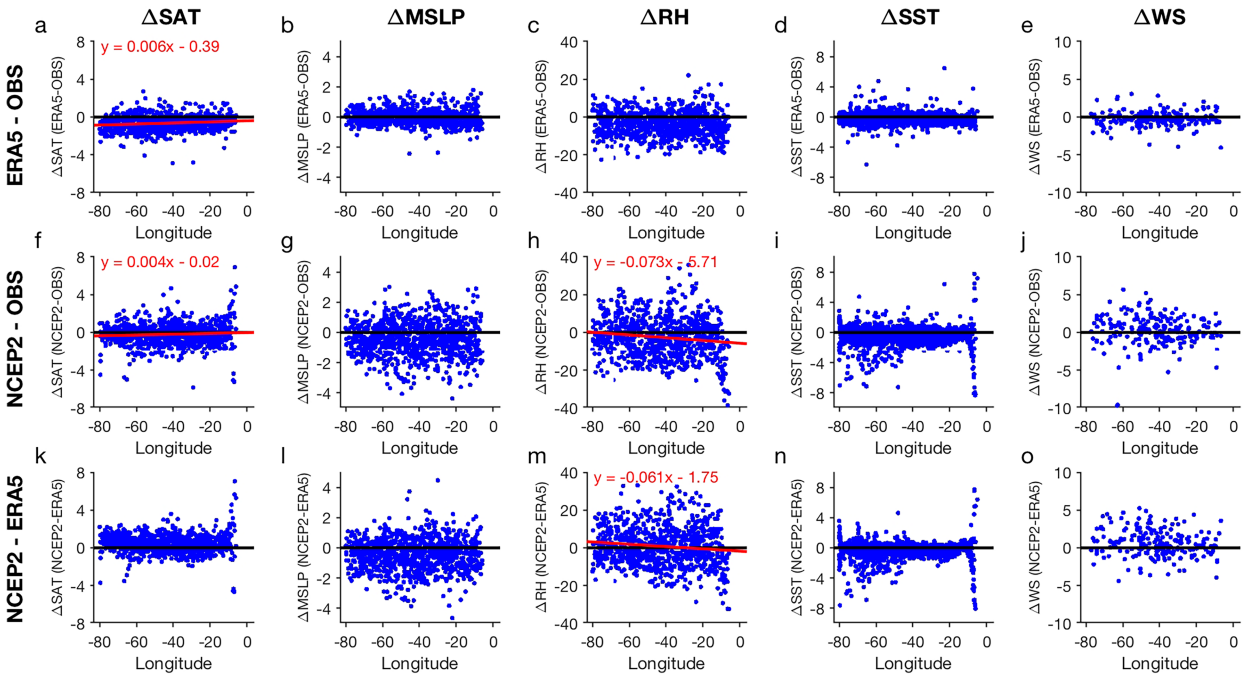

FIG. 6. Differences in the observed and reanalysis variables against longitude. The red lines represent linear trends, and only those with statistically significant dependence on longitude are shown. (first row) The differences between ERA5 and OBS variables, (second row) the differences between NCEP-2 and observed variables, and (third row) the differences between NCEP-2 and ERA5 variables. The variables and their units are as follows: SAT and SST (°C), MSLP (hPa), RH (%), and WS (m s^{-1}). For visibility and continuity, some graphs exclude large outliers.

accurately represent the full range of the observed SAT. Biases in ERA5 RH also show a weak dependence on the time of day (Fig. 5c), with the strongest negative bias during nighttime and the weakest negative bias during daytime. Biases in NCEP-2 RH (Fig. 5d) and biases in all other variables (not shown) did not show a clear dependence on the hour of the day.

In addition to biases in time of day, variable biases with respect to longitude were also assessed. Figure 6a indicates that the cold biases in ERA5 SAT are the strongest in the western North Atlantic Ocean, and the cold biases tend to decrease moving eastward. The regression of SAT difference against longitude (red line in Fig. 6a) suggests that the cold bias decreases from −0.85°C near the western boundary (80°W) to −0.42°C near the eastern boundary (5°W). The regression analysis also indicates a significant dependence of NCEP-2 SAT biases on longitude (Fig. 6f); however, this dependence is induced by the large warm biases near the eastern boundary. Those large outliers can also be seen in the differences between NCEP-2 and ERA5 SAT (Fig. 6k). After excluding data near the eastern boundary, east of 10°W, no significant dependence on longitude is found for biases in NCEP-2 SAT. Likewise, the longitudinal dependence of NCEP-2 RH biases (Figs. 6h,m) is also induced by the large outlier values near the eastern boundary. ERA5 RH tends to be lower than the observed RH throughout the domain without a significant dependence on longitude (Fig. 6c). Biases in ERA5 and NCEP-2 SST do not show a clear longitudinal dependence. However, similar to SAT and RH, NCEP-2 SST shows large biases near the eastern boundary. In the western North Atlantic (west of 40°W), we also see increased cold biases in NCEP-2 SST when compared to observations and ERA5 (Figs. 6i,n). The comparison of NCEP-2 SST with ERA5 SST (Fig. 6n) demonstrates a similar longitudinal distribution as its comparison with the observed SST (Fig. 6i).

No significant spatial trends were detected for biases in MSLP or WS, nor do they exhibit the same pattern of eastern North Atlantic outliers that SAT, RH, and SST do.

b. *Comparison of turbulent fluxes*

Here, we compare the turbulent (latent and sensible) heat fluxes derived from the measured variables and estimate the sensitivity of these fluxes on the variables from the scenarios outlined in section 3e. Key statistical characteristics of the THF estimates from all four scenarios and their comparisons are shown in Tables 1 and 2 and Fig. 7. Table 1 compares the mean, median, and standard deviation of THF. Regular numbers are derived from 103 data samples when fluxes from all four scenarios are available, and the bold numbers correspond to values derived from 837 data samples when fluxes from scenarios 2–4 are available. Table 2 shows the differences and correlations of THF estimates between different scenarios.

As shown in Table 1, LHF and SHF means for all scenarios are significantly larger than their respective medians. In general, the LHF medians are about 25 W m^{-2} lower than their corresponding means, except for scenario 4 showing a relatively lower

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

TABLE 1. Means, medians, and standard deviations of LHF and SHF estimates from all four scenarios (defined in section 3e). Regular numbers are derived from cases when flux estimates are available for all four scenarios (103 data points). Bold numbers are derived from cases when flux estimates are available for scenarios 2–4 (837 data points). All values are measured in watts per square meter. For all products, positive represents the direction into the atmosphere.

	LHF			SHF		
COARE flux scenario	Mean	Median	Std dev	Mean	Median	Std dev
1: OBS	93.67	70.75	89.64	7.60	2.94	17.09
2: OBS with ERA5 WS	87.27	61.96	83.78	6.81	2.73	16.03
	84.74	**61.42**	**79.98**	**6.76**	**2.83**	**18.71**
3: ERA5	100.26	65.93	98.01	6.38	1.74	15.08
	99.45	**76.00**	**88.30**	**7.05**	**2.19**	**18.78**
4: NCEP-2	96.27	78.23	71.08	6.08	1.19	14.64
	99.47	**85.13**	**69.71**	**5.57**	**1.39**	**19.52**

difference (<20 W m^{-2}) between its mean and median. SHF means for all scenarios are 4–5 W m^{-2} larger than their respective medians. Together, these suggest large outliers with high LHF and SHF in the subtropical North Atlantic Ocean during our study period. All standard deviations are also quite large relative to their means and medians, suggesting a wide range of values for all LHF and SHF scenarios. As expected from the large energy required during evaporation, the LHF means for all scenarios (~100 W m^{-2}) are much larger than the respective SHF means (<10 W m^{-2}).

Compared to the mean values of LHF and SHF computed exclusively from observations (scenario 1), both mean LHF and mean SHF from scenario 2 (all observations except ERA5 winds) are lower due to negative biases in ERA5 WS. Scenario 2 mean LHF is 6.40 W m^{-2} lower than scenario 1 mean LHF, and its mean SHF is 0.79 W m^{-2} lower than that from scenario 1. Considering individual cases, the majority of flux differences between scenarios 1 and 2 (90% of the total 103 data samples for LHF and 93% for SHF) are less than 5% of the flux itself. Flux estimates from scenarios 1 and 2 are also highly correlated with correlation coefficients of 0.99 for both LHF and SHF (Table 2).

Despite the lower biases in ERA5 WS, the mean LHF from scenario 3 (exclusively ERA5) is 6.59 W m^{-2} higher than that from scenario 1; this can be attributed to the negative biases in ERA5 RH and SAT. As can be seen from the comparison between scenarios 2 and 3, the biases in ERA5 RH and SST-SAT together induced a positive bias in LHF of 14.71 W m^{-2} (Tables 1 and 2; considering all 837 available data). The mean SHF from scenario 3 is lower than that from scenario 1, indicating that the SHF decrease induced by ERA5 WS bias is stronger than the SHF increase induced by ERA5 temperature biases. Similar to scenario 3, scenario 4 (exclusively NCEP-2) also presents higher mean LHF and lower mean SHF compared to scenario 1. However, the overestimate in NCEP-2 LHF is due to positive bias in NCEP-2 WS and negative bias in NCEP-2 RH, which is partially compensated for by the underestimate induced by cold bias in SST. For SHF, the cold bias in SST has a stronger impact than the positive bias in WS. Scenario 4 has the lowest variability in LHF. This low variability of NCEP-2 LHF (Table 1) is due to the fact that NCEP-2 does not capture the full range of observed RH; 93% of NCEP-2 RH values are within 60%–90% compared to the two-thirds of the observed RH falling in that range.

Although the mean flux values from scenarios 3 and 4 show more or less similar characteristics when compared with scenario 1, the correlation coefficients and the variability in the differences between scenarios 1 and 2 indicate that scenario 3 gives better flux estimates than scenario 4 (Table 2). ERA5 fluxes (scenario 3) are highly correlated with the observed

TABLE 2. Means, medians, and standard deviations of the differences between LHF and SHF from all four scenarios and their correlation coefficients. Flux labels adhere to the following convention: OBS–ERA5 (1–3) represents the fluxes calculated from ERA5 data subtracted from the fluxes calculated from observed data, with the numbers in parentheses showing the associated COARE scenario numbers. Fluxes are assessed in watts per meter square.

COARE scenario flux comparisons	Mean difference	Median difference	Std dev of differences	Correlation coefficient
LHF: OBS2–OBS (2–1)	−6.40	−3.96	13.65	0.99
SHF: OBS2–OBS (2–1)	−0.79	−0.25	2.58	0.99
LHF: ERA5–OBS (3–1)	6.59	2.55	36.90	0.93
SHF: ERA5–OBS (3–1)	−1.23	−2.20	8.19	0.88
LHF: NCEP-2–OBS (4–1)	2.60	0.29	59.56	0.75
SHF: NCEP-2–OBS (4–1)	−1.52	−1.59	11.53	0.77
LHF: ERA5–OBS2 (3–2)	14.71	8.50	34.76	0.92
SHF: ERA5–OBS2 (3–2)	0.29	−0.51	8.36	0.90
LHF: NCEP-2–OBS2 (4–2)	14.73	10.07	51.80	0.75
SHF: NCEP-2–OBS2 (4–2)	−1.19	−0.72	11.50	0.82
LHF: NCEP-2–ERA5 (4–3)	0.02	4.53	48.00	0.77
SHF: NCEP-2–ERA5 (4–3)	−1.48	−0.58	10.94	0.78

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

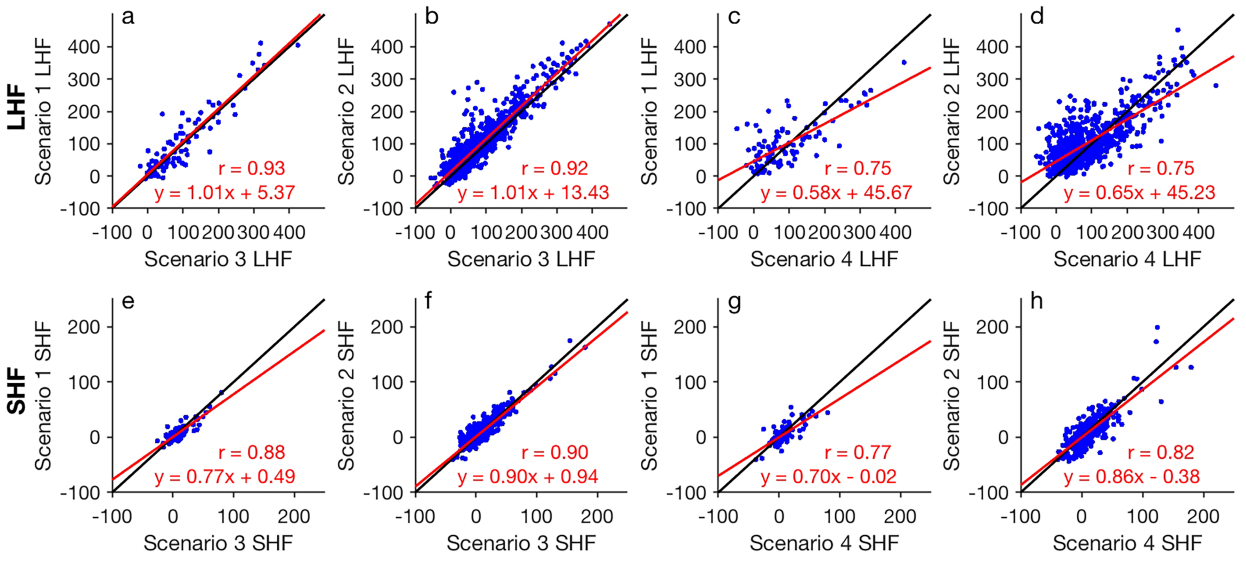

FIG. 7. Scatterplots comparing the (first row) LHF and (second row) SHF estimates. (a) Scenario 1 LHF against scenario 3 LHF, (b) scenario 2 LHF against scenario 3 LHF, (c) scenario 1 LHF against scenario 4 LHF, and (d) scenario 2 LHF against scenario 4 LHF. (e)–(h) As in (a)–(d), but for SHF. The black lines represent the 1:1 lines, and the red lines represent linear regression lines. All values are measured in watts per square meter. For visibility and continuity purposes, some graphs exclude large outliers. Note: scenario 1 (all observations), scenario 2 (observations with ERA5 winds), scenario 3 (ERA5), and scenario 4 (NCEP-2).

fluxes (scenarios 1 and 2), with correlation coefficients ranging between 0.88 and 0.93. NCEP-2 fluxes (scenario 4) are less correlated with the observed fluxes (scenarios 1 and 2), with correlation coefficients ranging between 0.75 and 0.82. The variabilities of flux differences between NCEP-2-based and observational-based estimates are considerably larger compared to the counterparts between ERA5-based and observation-based estimates (Table 2). For example, the variabilities of LHF differences between scenario 4 and scenarios 1 and 2 exceed 50 W m^{-2}, whereas the variabilities are around 35 W m^{-2} for the LHF differences between scenario 3 and scenarios 1 and 2. The large variability in the flux differences indicates that, despite the similar means in the flux estimates, the differences in individual cases can be large.

The better performance of ERA5-based estimates (scenario 3) can also be seen in the scatterplots comparing the LHF and SHF derived from different scenarios (Fig. 7). Both the regressions of LHF from scenarios 1 and 2 to scenario 3 give slopes close to 1.0 (Figs. 7a,b), whereas their regressions to scenario 4 LHF have slopes about 0.58–0.65 (Figs. 7c,d). The scatterplots of scenario 4 against scenarios 1 and 2 also show greater spread along the regression lines. The comparison for SHF shows similar results, with higher regression slopes and less spread between observed scenarios (1 and 2) and scenario 3 (Figs. 7e–h), though the differences are not as obvious as for LHF.

Similar to the variables, we examined the space and time dependence of SHF and LHF biases in scenarios 3 and 4 in comparison with scenario 2. Although not as apparent as for SAT biases, the SHF biases from both scenarios 3 and 4 show some dependence on the time of day (Figs. 8a,b), with the tendency of negative biases overnight and positive biases during daytime. This is consistent with stronger daytime cold biases in ERA5 and NCEP-2 SAT. LHF biases from both scenarios 3 and 4 do not show dependence on the time of day (Figs. 8c,d). For scenario 4, this agrees with the lack of time-of-day dependence in NCEP-2 RH biases. However, for scenario 3, the LHF biases are expected to depend on the time of day because of the diurnal changes in ERA5 RH biases (Fig. 5c). One possible explanation is that the effect of ERA5 SAT biases on LHF counteracts the effect from ERA5 RH biases. During the daytime, the potential larger positive LHF bias caused by stronger negative RH bias (enhanced evaporation) is compensated by the negative LHF bias induced by stronger negative SAT bias (reduced evaporation) and vice versa. No significant longitudinal dependence was found for all SHF and LHF biases (not shown) except the biases in scenario 3 SHF, which decrease from west to east because of the eastward-decreasing negative ERA5 SAT biases.

To better visualize the relationships of fluxes and the impacts of their key variables, Fig. 9 shows an example transect of the fluxes from scenarios 2–4 (omitting scenario 1 due to the limited availability of WS data) and key variables against longitude for the XBT transect carried out during July 2021. While the LHF and SHF of all depicted scenarios generally follow a similar longitudinal pattern, scenario 4 (NCEP-2) overestimates LHF across the entire section (Fig. 9a) because of the low biases in NCEP-2 RH (Fig. 9c) and stronger WS (Fig. 9e). Scenario 3 (ERA5) also overestimates the LHF in regions west of 55°W and near the eastern boundary, but to a much lesser degree. Both low biases in ERA5 RH and SAT contribute to the overestimation west of 55°W, whereas near the eastern boundary, the overestimation is mostly due to the low bias in ERA5 RH. The agreement among the three SHF estimates is much better. Scenario 4 SHF shows negative biases east of 20°W and between 70° and 60°W, while scenario 3 SHF has a positive bias near the western boundary and a

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

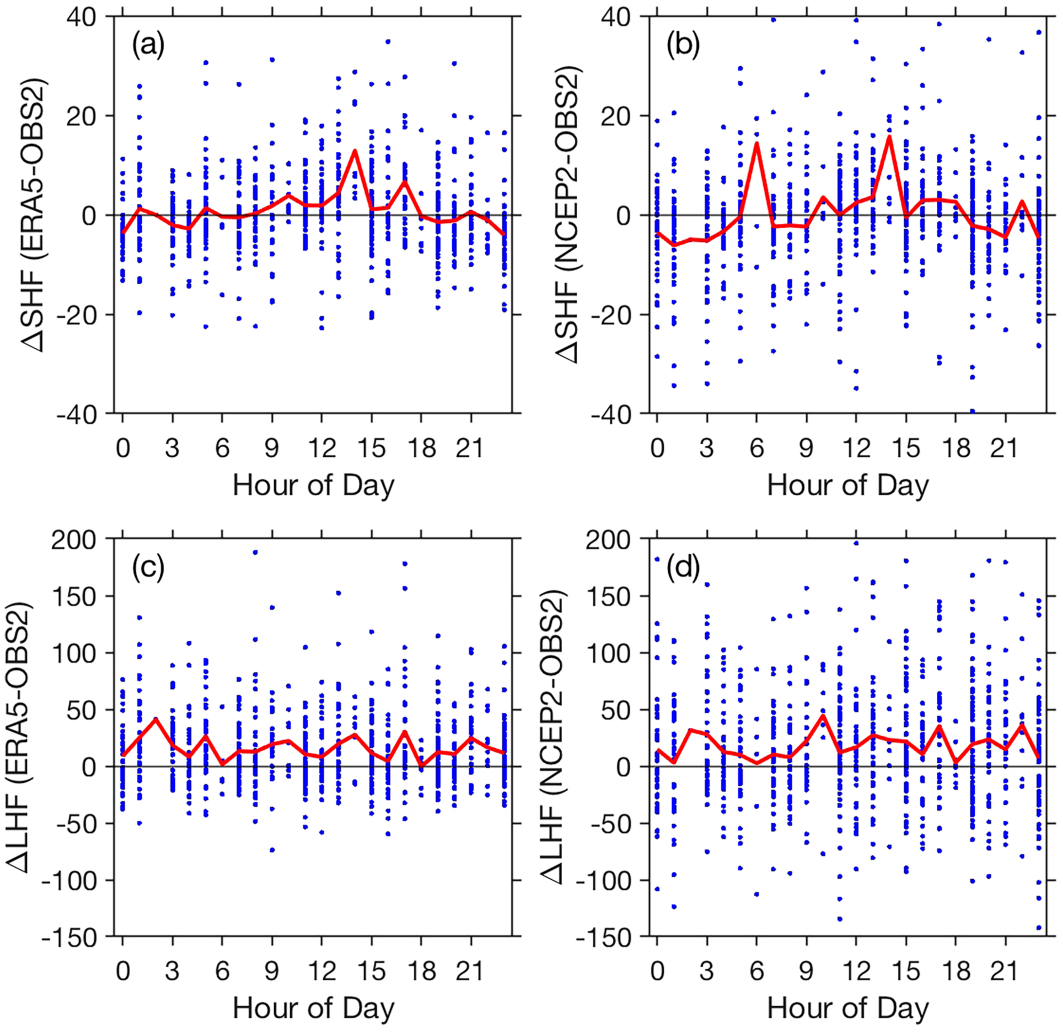

FIG. 8. Differences in SHF and LHF (W m^{-2}) between reanalysis-based and observation-based estimates against the hour of the day. (a) ΔSHF: ERA − OBS, (b) ΔSHF: NCEP-2 − OBS, (c) ΔLHF: ERA5 − OBS, and (d) ΔLHF: NCEP-2 − OBS. Here, OBS represents scenario 2 rather than scenario 1 for greater visibility. The red lines represent the mean difference for each hour.

negative bias near the eastern boundary (Fig. 9b). ERA5 SST agrees very well with the observed SST (Fig. 9f), and the biases in ERA5 SAT (Fig. 9d) are the dominant contributing factor for the biases in scenario 3 SHF. Both stronger NCEP-2 WS and negative biases in NCEP-2 SST-SAT contribute to the negative biases (stronger heat into the ocean) in scenario 3 SHF. Furthermore, both reanalysis products exhibit challenges in accurately reproducing the full range of observed values for fluxes and variables (visualized in the panels by the greater number of "spikes" in the scenario 2 data). This is especially evident in the SAT and SST variables (Figs. 9d,f), where NCEP-2 SAT and SST are particularly smooth, whereas the observed SAT and SST fluctuations are captured to some degree by ERA5 SAT and SST.

Figure 10 shows differences in fluxes between observed scenario 1 and reanalysis scenarios 3 and 4 against their respective key variables. As expected by formula 7, LHF biases are highly dependent on the differences in RH and to a lesser degree on the differences in WS. The trend lines and correlation coefficients show that 1) as the observed RH becomes larger than the reanalysis RH, the observed LHF becomes smaller than the reanalysis LHF (Figs. 10a,f), and 2) as the observed WS becomes larger than the reanalysis WS, the observed LHF becomes larger than the reanalysis LHF (Figs. 10b,g). Figure 10 also shows that SHF biases are highly dependent on the differences in SAT, and the dependence on the differences in SST and WS is rather weak. The trend lines and correlation coefficients show that 1) as the observed SAT becomes larger than the reanalysis SAT, the observed SHF becomes smaller than the reanalysis SHF; 2) as the observed SST becomes larger than the reanalysis SST, the observed SHF becomes larger than the reanalysis SHF; and 3) as the observed WS becomes larger than the reanalysis WS, the observed SHF becomes larger than the reanalysis SHF. These

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

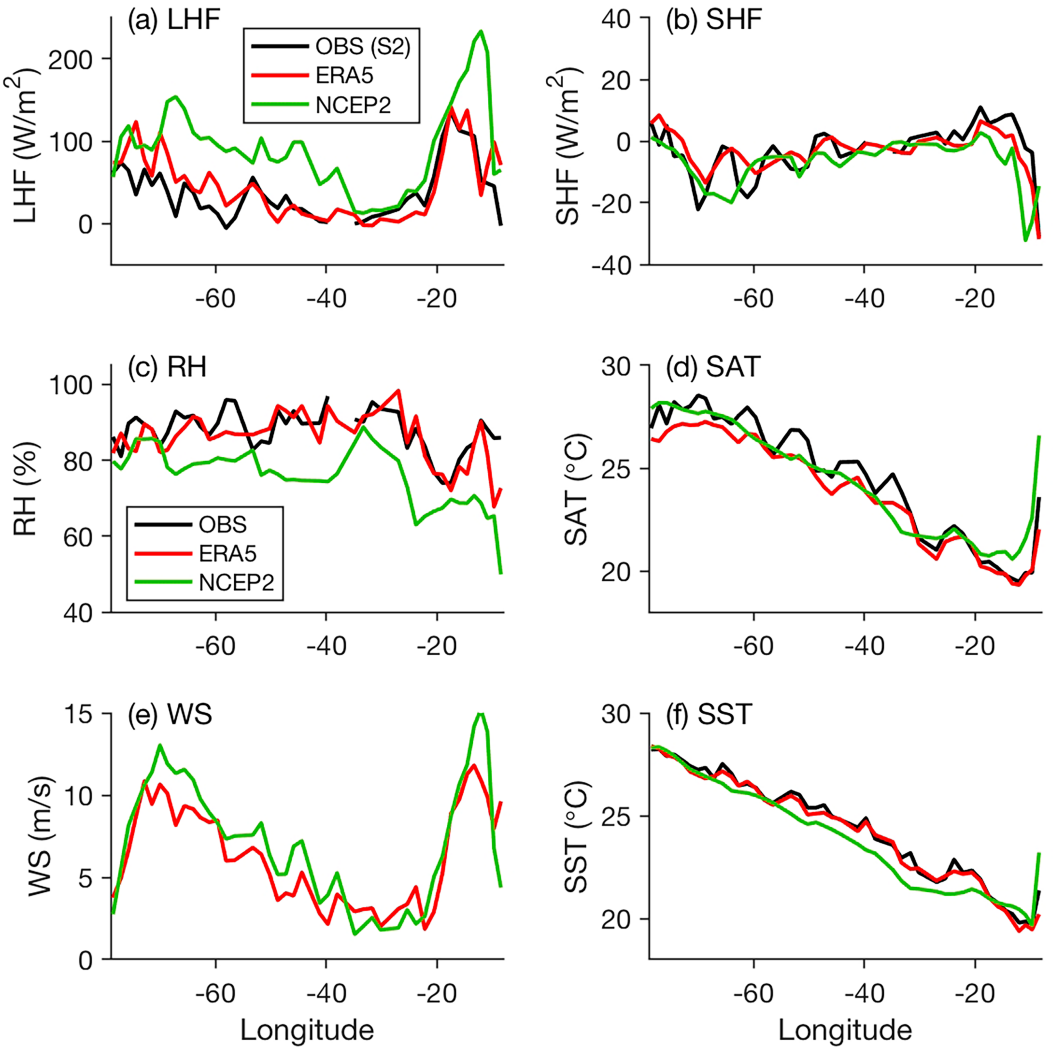

FIG. 9. Plots of variables against longitude for an example transect carried out in July 2021. (a) LHF and (b) SHF from scenarios 2 (black), 3 (red), and 4 (green) (scenario 1 is omitted due to considerably low data availability and therefore minimal visibility). (c) RH, (d) SAT, (e) WS, and (f) SST from observations (black), ERA5 (red), and NCEP-2 (green).

relationships correspond well to the sensible heat flux Eq. (8), where SHF is proportional to sea–air temperature difference and wind speed. Higher ocean temperature, thus larger temperature difference, and larger wind speed would increase SHF, whereas higher air temperature would reduce the temperature difference, thus decreasing SHF.

The same is true when comparing differences in reanalysis fluxes and variables with those used in scenario 2, which, although uses ERA5 WS, provides significantly more data output (Fig. 11). Figure 11 shows the same general trends found in Fig. 10. All slopes in Figs. 10 and 11 are of similar magnitude, except the slopes for ΔLHF against ΔSST between scenarios 2 and 3 (Fig. 11c) and for ΔSHF against ΔSAT between scenarios 2 and 4 (Fig. 11e) are somewhat lower. The correlation coefficients are also similar except the reduced correlation between ΔSHF and ΔSAT of scenarios 2 and 4. Figures 10 and 11 both show significantly higher spread when comparing observed and NCEP-2 fluxes and variables than when comparing observed and ERA5 fluxes and variables.

5. Discussion and conclusions

This study presents the first assessment of the value of NOAA/AOML's shipboard weather station data, which were launched in 2020 along XBT transects in the subtropical North Atlantic Ocean. Data collected from the weather stations and XBTs together with satellite measurements are used to assess biases and validity in the ERA5 and NCEP-2 reanalysis products with respect to turbulent heat fluxes. We calculated latent and sensible turbulent heat fluxes using the Coupled Ocean–Atmosphere Response Experiment (COARE) version 3.6 flux algorithm in four scenarios: 1) in situ data with satellite winds, 2) in situ data with ERA5 winds, 3) ERA5 data, and 4) NCEP-2 data. The resulting LHF and SHF from each scenario were

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

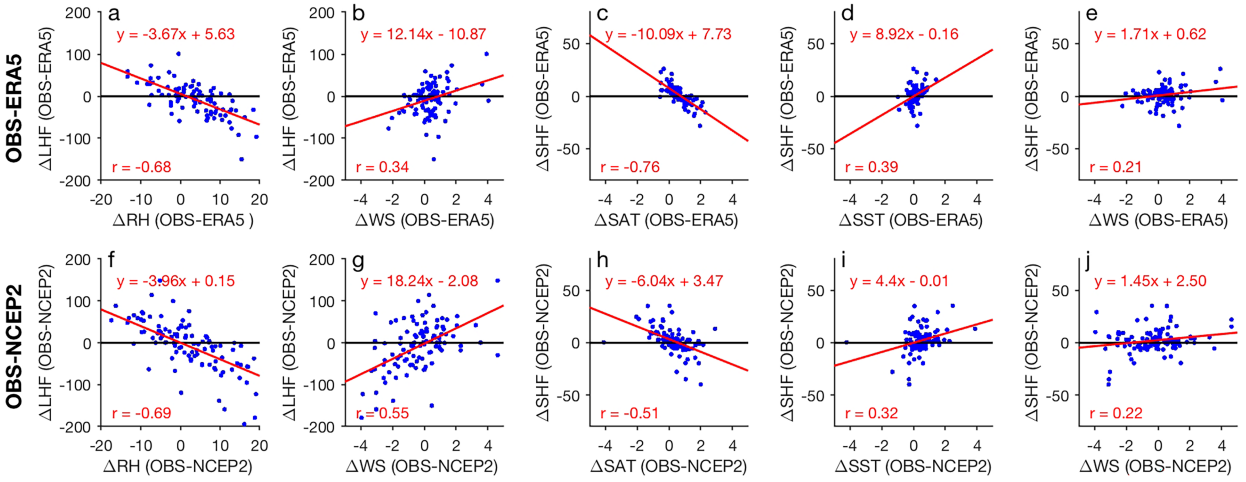

FIG. 10. Differences in flux estimates against the differences in their respective key variables. (top) Differences between scenarios 1 and 3, including differences in LHF (W m^{-2}) against differences in (a) RH (%) and (b) WS (m s^{-1}) and differences in SHF (W m^{-2}) against differences in (c) SAT (°C), (d) SST (°C), and (e) WS. (bottom) As in the top panels, but for differences between scenarios 1 and 4. For visibility and continuity, some graphs exclude large outliers.

evaluated with respect to key variables, time of day, and individual transects; each scenario was further compared with the results of the remaining scenarios to quantify correlations and mean and median differences.

Direct comparisons between variables of different products present four key results. First, across all products, SAT and SST tend to decrease toward the eastern Atlantic Ocean (Fig. 2). This is expected due to the warm water of the Gulf Stream in the west and the clockwise rotation of the subtropical gyre. Second, RH tends to disagree considerably across all three datasets. This is evident in Figs. 4 and 6, which show a comparatively weak correlation for all three RH products. ERA5 RH typically corresponds better with in situ RH than NCEP-2 RH does, although both relationships show consistently large differences between products. This disagreement can arise from multiple factors, such as the accuracy of the humidity sensors, which are

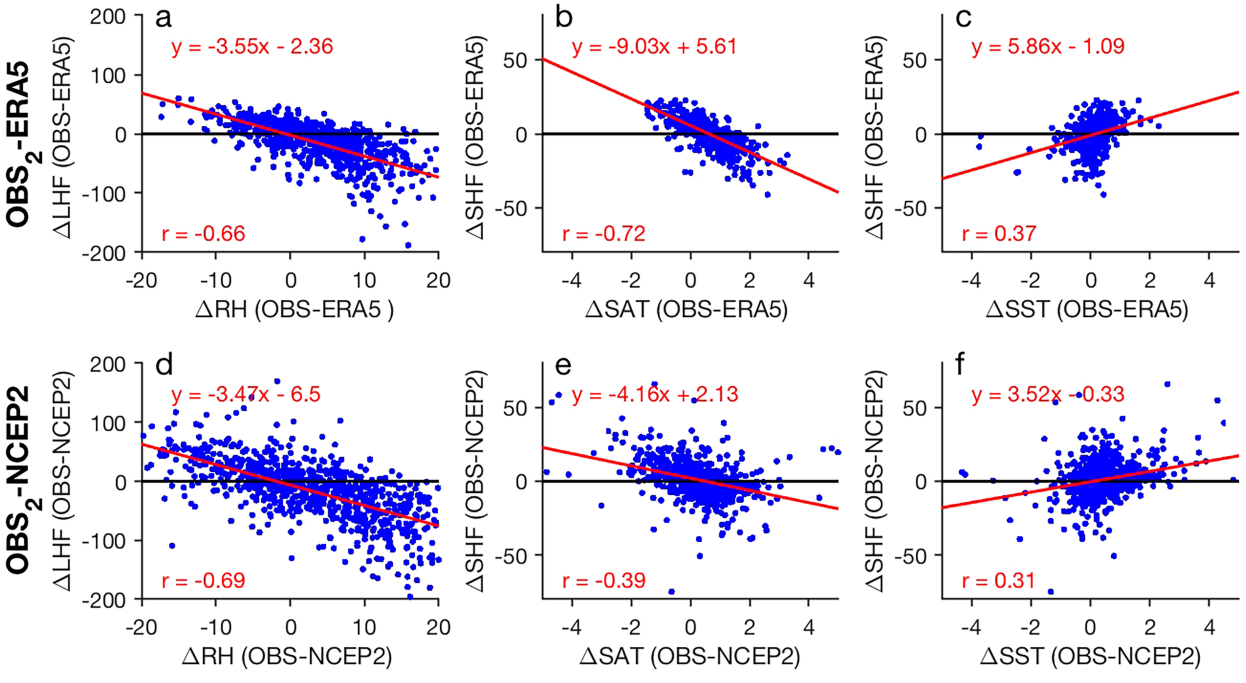

FIG. 11. As in Fig. 10, but comparing scenario 2 instead of scenario 1. (top) Differences between scenarios 2 and 3, including differences in LHF (W m^{-2}) against differences in (a) RH (%) and differences in SHF (W m^{-2}) against differences in (b) SAT and (c) SST (°C). (bottom) As in the top panels, but for differences between scenarios 2 and 4. WS is not included here due to scenario 2 utilizing ERA5 winds.

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

generally lower than the manufacturer's claims due to sensor quality and calibrations (Schellenberg 2002; Taylor and Cheng 2010), a larger spatial variability in RH than temperature (Hempel et al. 2018), high scatter at extreme RH values during cloudy and rainy days (Payne et al. 2002; Camuffo et al. 2017), and the effect of solar insolation on the measurements (McPhaden et al. 2009). Third, SAT from reanalysis products is highly correlated with in situ measurements, but it can have biases of a few degrees Celsius (Figs. 4 and 6). On average, both ERA5 and NCEP-2 SAT show a cold bias. For ERA5 SAT, this is predominantly present at higher temperatures in the western North Atlantic, and for NCEP-2 SAT, this is consistent across all temperatures. This cold bias is enhanced during the afternoon, which is in part due to the spatial and temporal resolution of reanalysis products (Fig. 5). Fourth, SST from reanalysis products is also highly correlated with in situ XBT measurements with a cold bias on average. In addition, biases between SST from XBT and reanalyses do not show a spatiotemporal dependence (Figs. 4 and 6), showing that in general their quality is robust, and XBT data in the top 5 m can be used as SST measurements. Although previous studies showed that the first few meters of the XBT profile are subject to transient signal for the thermistor to adapt to the surrounding water (Kizu and Hanawa 2002; Goes et al. 2017), the temperature in the top layers does not differ significantly from those in the surface mixed layer; thus, the quality of near-surface XBT data is a robust estimate of SST, except in regions where thermocline outcrops to surface or mixed layer depth is less than 5 m (Thadathil et al. 1999). This has yet to be tested with the present data.

These biases present challenges for the reanalysis flux calculations. Specifically, all reanalysis fluxes are limited by spatial and temporal smoothing; NCEP-2 data, which have a lower resolution in both space and time than ERA5 data, perform somewhat worse in comparisons with in situ data as well as with scenario 1 and 2 fluxes when run in the COARE 3.6 algorithm (Tables 1 and 2 and Figs. 4–11). The differences between NCEP-2 and in situ data are especially large and varied in the eastern Atlantic near Gibraltar, likely due to contamination of the coastal data with data over land, caused by the low spatial resolution of the product (Fig. 6). As expected in formulas 7–8, for both reanalysis products, biases in the LHF largely depend on the differences in RH and biases in SHF largely depend on the differences in SAT (Figs. 9–11).

Heat exchange between the atmosphere and ocean is an important process reflecting the interaction of the coupled climate system and affecting regional and global climate and weather (Trenberth and Fasullo 2010; Booth et al. 2017). Global air–sea heat fluxes have been used in numerous aspects of climate studies over many time scales and are needed to evaluate coupled atmosphere–ocean models and weather forecasting models (e.g., Cronin et al. 2019). Accurate estimates of air–sea heat fluxes are also important for evaluating regional and global heat budgets and for inferring or assessing oceanic meridional heat transports (e.g., Trenberth and Caron 2001; Grist and Josey 2003; Trenberth et al. 2019). However, commonly used reanalysis flux products can differ by 100 W m^{-2} or more due to biases in flux-related variables (Dong et al. 2010; Zhang et al. 2016; Pokhrel et al. 2020). VOS ships provide a great cost-effective platform to obtain simultaneous near-surface variables with good spatial coverage. Results presented in this study from analyzing data collected in the subtropical North Atlantic demonstrate the potential to improve turbulent flux estimates by reducing biases in near-surface atmospheric variables using VOS ships.

Acknowledgments. The authors thank NOAA-NWP and NOAA-GOMO for the funding, NOAA/AOML for the support and for the XBT network program, and the NOAA engineering group for developing, installing, and carrying out the meteorological observations. The authors also thank the Hapag-Lloyd Company for allowing the installation of weather stations on its container ships and the crew for deploying XBTs. This work was carried out under the University of Miami/CIMAS Cooperative Agreement NA20OAR4320472. The authors thank the William M. Lapenta NOAA Student Internship Program of summer 2022, which made this work possible.

Data availability statement. The ERA5 reanalysis was downloaded from the Copernicus Climate Change Service Climate Data Store (https://cds.climate.copernicus.eu/). The NCEP-2 reanalysis was downloaded from the NOAA Physical Sciences Laboratory (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html). The XBT data are available at the NOAA AOML webpage (https://www.aoml.noaa.gov/phod/hdenxbt/index.php). The satellite altimeter wind data are available at ftp://avisoftp.cnes.fr/AVISO/pub/ and ftp://ftp.star.nesdis.noaa.gov/pub/sod/lsa/cs2igdr/. The AODN portal can be accessed at https://portal.aodn.org.au/. Weather station data are available at https://www.aoml.noaa.gov/ftp/pub/phod/soopdata/XBT/.

REFERENCES

Bailey, R. J., H. E. Phillips, and G. Meyers, 1989: Relevance to TOGA of systematic XBT errors. *Proceedings of the Western Pacific International Meeting and Workshop on TOGA COARE*, J. Picaut et al., Eds., Ostrom, Noumea, New Caledonia, 775–784.

Bellucci, A., and Coauthors, 2021: Air-sea interaction over the Gulf Stream in an ensemble of HighResMIP present climate simulations. *Climate Dyn.*, **56**, 2093–2111, https://doi.org/10.1007/s00382-020-05573-z.

Bishop, S. P., R. J. Small, F. O. Bryan, and R. A. Tomas, 2017: Scale dependence of midlatitude air–sea interaction. *J. Climate*, **30**, 8207–8221, https://doi.org/10.1175/JCLI-D-17-0159.1.

Blanc, T. V., 1987: Accuracy of bulk-method-determined flux, stability, and sea surface roughness. *J. Geophys. Res.*, **92**, 3867–3876, https://doi.org/10.1029/JC092iC04p03867.

Booth, J. F., Y.-O. Kwon, S. Ko, R. J. Small, and R. Msadek, 2017: Spatial patterns and intensity of the surface storm tracks in CMIP5 models. *J. Climate*, **30**, 4965–4981, https://doi.org/10.1175/JCLI-D-16-0228.1.

Bourassa, M. A., D. M. Legler, J. J. O'Brien, and S. R. Smith, 2003: SeaWinds validation with research vessels. *J. Geophys. Res.*, **108**, 3019, https://doi.org/10.1029/2001JC001028.

Brodeau, L., B. Barnier, S. K. Gulev, and C. Woods, 2017: Climatologically significant effects of some approximations in the

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

bulk parameterizations of turbulent air–sea fluxes. *J. Phys. Oceanogr.*, **47**, 5–28, https://doi.org/10.1175/JPO-D-16-0169.1.

Brunke, M. A., X. Zeng, and S. Anderson, 2002: Uncertainties in sea surface turbulent flux algorithms and data sets. *J. Geophys. Res.*, **107**, 3141, https://doi.org/10.1029/2001JC000992.

Campos, R. M., 2023: Analysis of spatial and temporal criteria for altimeter collocation of significant wave height and wind speed data in deep waters. *Remote Sens.*, **15**, 2203, https://doi.org/10.3390/rs15082203.

——, J.-H. G. M. Alves, S. G. Penny, and V. Krasnopolsky, 2020: Global assessments of the NCEP Ensemble Forecast System using altimeter data. *Ocean Dyn.*, **70**, 405–419, https://doi.org/10.1007/s10236-019-01329-4.

——, C. B. Gramcianinov, R. de Camargo, and P. L. da Silva Dias, 2022: Assessment and calibration of ERA5 severe winds in the Atlantic Ocean using satellite data. *Remote Sens.*, **14**, 4918, https://doi.org/10.3390/rs14194918.

Camuffo, D., A. della Valle, and F. Becherini, 2017: A critical analysis of one standard and five methods to monitor surface wetness and time-of-wetness. *Theor. Appl. Climatol.*, **132**, 1143–1151, https://doi.org/10.1007/s00704-017-2167-9.

Cheng, L., J. Zhu, R. Cowley, T. Boyer, and S. Wijffels, 2014: Time, probe type, and temperature variable bias corrections to historical expendable bathythermograph observations. *J. Atmos. Oceanic Technol.*, **31**, 1793–1825, https://doi.org/10.1175/JTECH-D-13-00197.1.

Cronin, M. F., and Coauthors, 2019: Air-sea fluxes with a focus on heat and momentum. *Front. Mar. Sci.*, **6**, 430, https://doi.org/10.3389/fmars.2019.00430.

Dong, S., S. T. Gille, J. Sprintall, and E. J. Fetzer, 2010: Assessing the potential of the Atmospheric Infrared Sounder (AIRS) surface temperature and specific humidity in turbulent heat flux estimates in the Southern Ocean. *J. Geophys. Res.*, **115**, C05013, https://doi.org/10.1029/2009JC005542.

Edson, J. B., and Coauthors, 2013: On the exchange of momentum over the open ocean. *J. Phys. Oceanogr.*, **43**, 1589–1610, https://doi.org/10.1175/JPO-D-12-0173.1.

Eymard, L., and Coauthors, 1999: Surface fluxes in the North Atlantic current during CATCH/FASTEX. *Quart. J. Roy. Meteor. Soc.*, **125**, 3563–3599, https://doi.org/10.1002/qj.49712556121.

Fairall, C. W., E. F. Bradley, J. E. Hare, A. A. Grachev, and J. B. Edson, 2003: Bulk parameterization of air–sea fluxes: Updates and verification for the COARE algorithm. *J. Climate*, **16**, 571–591, https://doi.org/10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.

——, and Coauthors, 2011: Implementation of the Coupled Ocean-Atmosphere Response Experiment flux algorithm with CO_2, dimethyl sulfide, and O_3. *J. Geophys. Res.*, **116**, C00F09, https://doi.org/10.1029/2010JC006884.

Frankignoul, C., and K. Hasselmann, 1977: Stochastic climate models, Part II Application to sea-surface temperature anomalies and thermocline variability. *Tellus*, **29A**, 289–305, https://doi.org/10.3402/tellusa.v29i4.11362.

Goes, M., E. Babcock, F. Bringas, P. Ortner, and G. Goni, 2017: The impact of improved thermistor calibration on the expendable bathythermograph profile data. *J. Atmos. Oceanic Technol.*, **34**, 1947–1961, https://doi.org/10.1175/JTECH-D-17-0024.1.

Goni, G. J., and Coauthors, 2019: More than 50 years of successful continuous temperature section measurements by the global expendable bathythermograph network, its integrability, societal benefits, and future. *Front. Mar. Sci.*, **6**, 452, https://doi.org/10.3389/fmars.2019.00452.

Grist, J. P., and S. A. Josey, 2003: Inverse analysis adjustment of the SOC air–sea flux climatology using ocean heat transport constraints. *J. Climate*, **16**, 3274–3295, https://doi.org/10.1175/1520-0442(2003)016<3274:IAAOTS>2.0.CO;2.

Hempel, S., M. König, C. Menz, D. Janke, B. Amon, T. M. Banhazi, F. Estellés, and T. Amon, 2018: Uncertainty in the measurement of indoor temperature and humidity in naturally ventilated dairy buildings as influenced by measurement technique and data variability. *Biosyst. Eng.*, **166**, 58–75, https://doi.org/10.1016/j.biosystemseng.2017.11.004.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hoffmann, L., and Coauthors, 2019: From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmos. Chem. Phys.*, **19**, 3097–3124, https://doi.org/10.5194/acp-19-3097-2019.

Jin, X., and L. Yu, 2013: Assessing high-resolution analysis of surface heat fluxes in the Gulf Stream region. *J. Geophys. Res. Oceans*, **118**, 5353–5375, https://doi.org/10.1002/jgrc.20386.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, https://doi.org/10.1175/BAMS-83-11-1631.

Kelly, K. A., R. J. Small, R. M. Samelson, B. Qiu, T. M. Joyce, Y.-O. Kwon, and M. F. Cronin, 2010: Western boundary currents and frontal air–sea interaction: Gulf Stream and Kuroshio Extension. *J. Climate*, **23**, 5644–5667, https://doi.org/10.1175/2010JCLI3346.1.

Kizu, S., and K. Hanawa, 2002: Start-up transient of XBT measurement. *Deep-Sea Res. I*, **49**, 935–940, https://doi.org/10.1016/S0967-0637(02)00003-1.

Laurindo, L. C., A. J. Mariano, and R. Lumpkin, 2017: An improved near-surface velocity climatology for the global ocean from drifter observations. *Deep-Sea Res. I*, **124**, 73–92, https://doi.org/10.1016/j.dsr.2017.04.009.

McPhaden, M. J., and Coauthors, 2009: RAMA: The research moored array for African–Asian–Australian monsoon analysis and prediction. *Bull. Amer. Meteor. Soc.*, **90**, 459–480, https://doi.org/10.1175/2008BAMS2608.1.

Moore, G. W. K., and I. A. Renfrew, 2002: An assessment of the surface turbulent heat fluxes from the NCEP–NCAR reanalysis over the western boundary currents. *J. Climate*, **15**, 2020–2037, https://doi.org/10.1175/1520-0442(2002)015<2020:AAOTST>2.0.CO;2.

Payne, R. E., and Coauthors, 2002: A comparison of buoy meteorological systems. UOP Technical Report 02-05, Woods Hole Oceanographic Institution, 67 pp.

Pokhrel, S., U. Dutta, H. Rahaman, H. Chaudhari, A. Hazra, S. K. Saha, and C. Veeranjaneyulu, 2020: Evaluation of different heat flux products over the tropical Indian Ocean. *Earth Space Sci.*, **7**, e2019EA000988, https://doi.org/10.1029/2019EA000988.

Ribal, A., and I. R. Young, 2019: 33 years of globally calibrated wave height and wind speed data based on altimeter observations. *Sci. Data*, **6**, 77, https://doi.org/10.1038/s41597-019-0083-9.

——, and ——, 2020: Calibration and cross validation of global ocean wind speed based on scatterometer observation. *J. Atmos. Oceanic Technol.*, **37**, 279–297, https://doi.org/10.1175/JTECH-D-19-0119.1.

Schellenberg, R., 2002: The Trouble with humidity: the hidden challenges of RH calibration. Veriteq instruments, 3 pp.,

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC

http://www.vaisala.com/Vaisala%20Documents/White%20Papers/lsh-Trouble-with-Humidity.pdf.

Smith, S. R., D. M. Legler, and K. V. Verzone, 2001: Quantifying uncertainties in NCEP reanalyses using high-quality research vessel observations. *J. Climate*, **14**, 4062–4072, https://doi.org/10.1175/1520-0442(2001)014<4062:QUINRU>2.0.CO;2.

——, K. Briggs, M. A. Bourassa, J. Elya, and C. R. Paver, 2018: Shipboard automated meteorological and oceanographic system data archive: 2005–2017. *Geosci. Data J.*, **6**, 73–86, https://doi.org/10.1002/gdj3.59.

Stull, R. B., 2017: *Practical Meteorology: An Algebra-based Survey of Atmospheric Science*. Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, 926 pp.

Taylor, S. T., and C. H. Cheng, 2010: Economizer high limit controls and why enthalpy economizers don't work. *ASHRAE J.*, **52**, 12–28.

Thadathil, P., A. K. Ghosh, J. Pattanaik, and J. Ratnakaran, 1999: A quality-control procedure for surface temperature and surface layer inversion in the XBT data archive from the Indian Ocean. *J. Atmos. Oceanic Technol.*, **16**, 980–982, https://doi.org/10.1175/1520-0426(1999)016<0980:AQCPFS>2.0.CO;2.

Trenberth, K. E., and J. M. Caron, 2001: Estimates of meridional atmosphere and ocean heat transports. *J. Climate*, **14**, 3433–3443, https://doi.org/10.1175/1520-0442(2001)014<3433:EOMAAO>2.0.CO;2.

——, and J. T. Fasullo, 2010: Tracking Earth's energy. *Science*, **328**, 316–317, https://doi.org/10.1126/science.1187272.

——, Y. Zhang, J. T. Fasullo, and L. Cheng, 2019: Observation-based estimates of global and basin ocean meridional heat transport time series. *J. Climate*, **32**, 4567–4583, https://doi.org/10.1175/JCLI-D-18-0872.1.

Valdivieso, M., and Coauthors, 2017: An assessment of air–sea heat fluxes from ocean and coupled reanalyses. *Climate Dyn.*, **49**, 983–1008, https://doi.org/10.1007/s00382-015-2843-3.

Wu, R., B. P. Kirtman, and K. Pegion, 2006: Local air–sea relationship in observations and model simulations. *J. Climate*, **19**, 4914–4932, https://doi.org/10.1175/JCLI3904.1.

Zhang, D., M. F. Cronin, C. Wen, Y. Xue, A. Kumar, and D. McClurg, 2016: Assessing surface heat fluxes in atmospheric reanalyses with a decade of data from the NOAA Kuroshio Extension Observatory. *J. Geophys. Res. Oceans*, **121**, 6874–6890, https://doi.org/10.1002/2016JC011905.

Brought to you by NOAA Library | Unauthenticated | Downloaded 04/01/25 06:31 PM UTC