Ocean Reference Stations: Long-term in situ observations of surface meteorology and air-sea fluxes at fixed open ocean locations are essential

Dr. Robert A Weller\*, Woods Hole Oceanographic Institution, Woods Hole MA, 02543

Dr. Roger Lukas and Dr. James Potemra, University of Hawaii at Manoa, Honolulu, HI 96822

Dr. Albert. J. Plueddemann, Woods Hole Oceanographic Institution, Woods Hole MA, 02543

Dr. Chris Fairall, NOAA Physical Sciences Laboratory, Boulder, CO 80305

Dr. Sebastien Bigorre, Woods Hole Oceanographic Institution, Woods Hole MA, 02543

\*Corresponding author rweller@whoi.edu

October 19, 2021

Abstract:

At three extra-equatorial open ocean sites in distinct tradewind regimes of the Pacific and Atlantic surface moorings first deployed in the early 2000s are maintained to collect accurate near-surface atmospheric and upper ocean measurements, to capture long-running records of oceanic and atmospheric variability and of coupling between the atmosphere and ocean, and to provide benchmarks for assessing models. Wind speed and direction, air temperature and humidity, rainfall, downward broadband shortwave and longwave irradiances, and barometric pressure are measured every minute. Together with sea surface temperature (SST) and near-surface current these are used to estimate the air-sea exchanges of heat, freshwater, and momentum. Laboratory calibrations, intercomparisons of redundant sensors on the moorings, analysis of overlapping mooring data, and field comparisons with shipboard sensors support assessment of the quality of the observations and derived air-sea fluxes. These surface meteorological observations are withheld from near real time incorporation into atmospheric models and remote sensing products, providing an independent means to assess models and remote sensing products. These three sites are therefore designated as Ocean Reference Stations (ORS). Comparisons are made between ORS time series and both modern reanalyses (ERA5, MERRA2, NCEP2) and CMIP6 models to gain insight into the models' ability to represent low-frequency variability and multi-year means. Biases in air-sea heat flux components from reanalyses are significant. Most CMIP6 SST values are significantly biased relative to the moorings, some by more than 3.0°C; and reanalyses and CMIP6 models fail to provide as much heat into the ocean as observed.

Capsule:

Careful measurements made from three surface buoys in the Atlantic and Pacific tradewind regions are ongoing and now provide close to 20-year long time series of surface meteorology and atmosphere-ocean exchanges of heat, freshwater, and momentum. These data are not used in models, and these independent benchmark time series are compared to modern reanalyses and climate models.

**1. Introduction**

In the open ocean, away from coasts and the equator, there are few sites where continuous, long-term, time-resolved observations in the atmospheric and oceanic surface layers are collected. Surface buoys, moored to the seafloor, provide unique platforms for meteorological sensors. Careful cross-calibrations of instruments on successive deployments enables merged long, accurate, and consistent time series with high resolution (once per minute or better). These long time series increase our understanding of the role of the ocean in the earth's weather and climate and serve as benchmarks to assess the realism of reanalyses and coupled models.

This focus here is on the surface observations and derived air-sea fluxes from three sites in distinct extra-equatorial tradewind regions (Figure 1). Tradewinds cover ~50% of the world ocean, flowing westward and equatorward in the lower portion of regional Hadley Circulations, outward from subsidence in subtropical high-pressure centers into ascending deep convection in the inter-tropical convergence zones (ITCZs). The Stratus Ocean Reference Station (ORS) site, 1,500 km west of northern Chile, where marine stratocumulus is characteristic, was established in October 2000. The Northwest Tropical Atlantic Station (NTAS) ORS was established in March 2001 1,100 km east of the island of Martinique in the Caribbean. The Woods Hole Oceanographic Institution (WHOI) Hawaii Ocean Time-series (HOT) Station, known as the WHOTS ORS, was first deployed in August 2004 at Station ALOHA 100 km north of Oahu.

The observing goals of the ORS effort are to collect surface meteorological and upper ocean data with high reliability and accuracy, to carry out calibrations and at-sea intercomparisons necessary to support quality control assessments, and to make merged climate data records with documented accuracy estimates widely available. As the records are extended, they provide unique and important information about temporal variability, opening opportunities to study oceanic, atmospheric, and coupled variability on intra-annual, decadal, and, eventually, multi-decadal time scales. They also continue to serve as focal points for process studies. Because the ORS observations are withheld from assimilation into initial conditions for numerical weather prediction models and subsequent reanalysis products, and because they are not incorporated into remote sensing products, these time series facilitate independent quantitative assessment of surface meteorology and air-sea fluxes from those reanalyses [e.g. Valdivieso et al. (2017)], from coupled climate models [e.g. Zuidema et al., (2016)], from remote sensing products [e.g. Pinker et al. (2018)], and from hybrid combinations of model and remote sensing fields [e.g. Yu and Weller, (2007)]. As such they are an essential element of our observing and modeling efforts.

In this report, we use merged ORS time series to gain insights into how well three modern reanalyses and recent coupled models represent low frequency and mean sea surface temperature (SST), wind stress and air-sea heat flux. The technical approach to collecting the observations is briefly summarized in Section 2 and in more detail in Weller et al. (2012) and Weller (2016). Section 3 summarizes assessments of the quality of the surface meteorological observations and derived air-sea fluxes. Quantifying the uncertainties associated with the ORS time series sets the stage their use as a benchmarks. Section 4 presents an overview of the surface meteorology and air-sea fluxes at each of the three ORS. Section 5 focuses on low pass filtered time series and annual and longer term means and comparisons between ORS time series and modern reanalyses and climate models. Section 6 presents discussions and conclusions.

**2. Equipping and Maintaining the Ocean Reference Stations**

In recent decades surface mooring technology has improved (Davis et al. 2018), and surface moorings deployed equatorial waters in the Pacific Ocean (McPhaden 1995), the Atlantic Ocean (Bourlès et al. 2019) and the Indian Ocean (McPhaden et al. 2009) are key components of sustained ocean observing. Away from the equator, conditions are more challenging and the three ORS are among the few long-term surface moorings maintained in the open ocean outside the tropics. Internationally, long-term moorings, including the ORS, are coordinated by OceanSITES (http://www.oceansites.org) as part of the Global Ocean Observing System (GOOS). Efforts in the United States to build a network of ocean reference sites are supported by the National Oceanic and Atmospheric Administration (https://globalocean.noaa.gov/Research/Ocean-Reference-Stations-OceanSITES).

Figure 2 shows the towers of both the initial 3-m aluminum ORS buoy hull and the more recent closed cell foam 3-m hull. A wind vane bolted to the tower orients the buoy and provides an upwind face for mounting anemometers, air temperature and air humidity sensors. Selection of meteorological sensors sought accuracy; reliability; redundancy; low power consumption; low susceptibility to corrosion, radio frequency interference, and salt spray; internal storage of key metadata (sensor model and serial number, calibration information); easy digital communication, and ability to deploy sensors either as stand-alone units or linked together to form a system. To meet these requirements, two separate modular systems, originally called the IMET (Improved METeorological) system and later the ASIMET (Air-Sea Interaction METeorological) system (Hosom et al., 1995), are deployed sampling once per minute running for over a year on battery power. ASIMET modules measure downward shortwave radiation (DSWR), downward longwave radiation (DLWR), air temperature (Ta) and humidity (qa), accumulated precipitation (P), barometric pressure (SLP), wind speed (WSPD) and direction (WDIR). Radiometers are mounted above all other structures to avoid shadows. Anemometers, Ta and qa sensors are mounted on the upwind face of the tower. P and SLP modules are mounted in the tower center. Modules measuring ocean temperature and salinity are mounted ~1.8 m below the surface. The sensors in current use and their mean heights/depths are given in Table 1. Since 2010 systems investigators at NOAA Pacific Marine Environmental Laboratory (PMEL) have installed on Stratus and WHOTS have measured the partial pressures of carbon dioxide (pCO2) above and below the surface (Sutton et al., 2014).

Recovery of a deployed surface mooring, preceded by nearby deployment of a fresh surface mooring, is planned annually. After recovery, calibrated and quality-controlled data are used together with the COARE bulk formulae (Fairall et al., 1996, 2003) to compute heat, freshwater, and momentum fluxes. Further analysis and processing are carried out to merge successive deployments as climate data records. As the bulk formulae are updated or when new information is available about calibration or processing procedures, the deployment-by-deployment data are reprocessed and updated merged time series created. During each deployment, hourly surface meteorological data are telemetered to support checking ORS systems; telemetry also aids testing on land and at sea comparisons between the ORS and shipboard sensors. The telemetered data are, however, purposefully not made available to the Global Telecommunication System (GTS) in order to keep the ORS time series as independent benchmarks.

Fifty-four of 58 planned annual cruises have been conducted as planned. The research vessel is equipped with an additional set of surface meteorological sensors. For many cruises, these additional sensors have come from the NOAA Physical Sciences Laboratory, Boulder, CO (Fairall et al., 2008). If NOAA PSL is not participating, additional freshly calibrated stand-alone ASIMET modules are installed on the ship. Upon arrival at the ORS site, the fresh surface mooring is deployed and one to several days of data are collected with both old and new ORS buoys in the water. During that overlap time the ship spends time on station downwind of the buoy with bow into the wind. After the *in-situ* intercomparison the old ORS mooring is recovered. Recovered ASIMET modules are photographed, and data spikes induced (for example, by putting covers on the shortwave radiometers and placing SST sensors in ice water) to allow checking and refining the time bases of the sensors. The returned ASIMET modules are post-calibrated; any performance anomalies during the deployments cause modules to be flagged for closer scrutiny. Overall data return has been high, but occasional failures of the mooring have resulted in shorter-than-planned deployments. Total gap time at Stratus is 1.8 years out of 21 years planned. Total gap time at NTAS is 2.4 years out of 20 years planned. There are no gaps in the occupancy of the WHOTS ORS site.

**3. The quality of ORS surface meteorological observations and derived air-sea fluxes**

ORS meteorological sensors are exposed to solar heating, salt spray, bird guano, birds attempting to land, and contact with ships. After quality checks, data from the two redundant ASIMET systems and additional stand-alone sensors have been able to be merged to yield a complete one-minute data set 99.9% of the time the ORS have been on station. In addition to the goal of obtaining complete time series, a focus has been on quantifying accuracies. Colbo and Weller (2000) assessed the accuracy of the IMET sensors and used error propagation equations derived from the COARE 3.0 bulk formulae (Fairall et al., 1996, 2003) to estimate the accuracy of the fluxes. Differences between redundant one-minute samples also occur from clock drift. Some errors, such as from radiative heating of passively ventilated sensors, buoy tilt, salt spray, flow distortion, and anemometer bearing friction, depend on wind speed. Table 2 summarizes the field accuracies of the IMET sensors. The accuracy of the self-siphoning rain gauge results from under-catchment (Nespor and Sevruk, 1999). Table 3 summarizes the accuracies of annual averages of the bulk formulae fluxes. Some errors in the individual net heat flux component measurements offset each other resulting in a smaller net heat flux error than the sum of the errors of each component. In light winds, for example, radiative heating errors impact sensors but buoy tilt-related errors are smaller, while for stronger winds radiative heating errors are small but tilt-related errors may increase.

**4. An overview of surface meteorology and air-sea fluxes at the ORS**

Figure 3 shows daily averages of the time series of surface meteorology at Stratus; matching figures for WHOTS and NTAS are provided in the Supplementary Material. At all three ORS annual cycles are evident in DSWR, SST, Ta, qa, net air-sea heat flux (Qnet), SLP and net longwave radiation (QLW). To avoid potential biases and false trends, long-term means were computed using only full calendar years. WHOTS has 14 full years, Stratus 15 full years, and NTAS 16 full years. The means of the surface meteorological variables over these years are in Table 4.

Some of the differences can be understood within the Lagrangian framework of the tradewind boundary layer (Albrecht, 1979, 1984; Albrecht et al., 1995 a, b; Bretherton et al., 1999; Albrecht et al., 2019), recognizing that the ORS sample key portions of the surface easterly tradewind branch of the regional Hadley circulations. Stratus is in the strong subsidence of the subtropical high of the eastern basin with associated stratocumulus deck. WHOTS is in the fair-weather cumulus core of the tradewinds, equatorward and west of the subtropical high where the surface layer becomes decoupled from the troposphere. NTAS is in the region of strong moisture convergence close to deep tropical convection in the western tropics poleward of the ITCZ. Translating Stratus to the eastern North Pacific, the corresponding virtual site is 32°N 125°W, further from the equator than Stratus because of the northern hemisphere location of the ITCZ, and a comparable distance west of the continent. This virtual time-series must be shifted by six months from the observed Stratus record to account for the seasonal phasing, and the winds rotated 90° counterclockwise. Translating NTAS from the Atlantic to a virtual site in the Pacific based on mean SST and proximity to the summertime ITCZ puts it around 12°N, 170°E.

The long-term mean DSWR is smallest at Stratus (205.2 W m-2) and largest at NTAS (246.7 W m-2) due to decreasing latitude and cloud fraction downstream. One of the findings at the Stratus ORS is much greater day-to-day variability in cloud cover than anticipated, including periods of clear skies, of low marine stratus, and of mixed stratus and cumulus. Mean DLWR increases downstream from Stratus to WHOTS to NTAS due to warming, moistening, and thickening of the planetary boundary layer (PBL). NTAS, with the warmest mean Ta and highest qa, has the largest mean DLWR (405.9 W m-2) compared to WHOTS (388.8 W m-2) and Stratus (376.8 W m-2). Mean SSTs were 20.3°C, 25.1° and 27.2°C at Stratus, WHOTS, and NTAS, consistent with greater ocean absorption of SWR downstream. One-day SST maxima over the full mean daily time series of 28.7°C at WHOTS and 29.7°C at NTAS were above the ~26°C threshold for deep convection, but only reached 24.7°C at Stratus. Stratus has the coolest Ta and NTAS the warmest, due to downstream heating from the ocean below, and reduced entrainment above the PBL, with qa also increasing accordingly. Mean SST was 0.9°C warmer than Ta at Stratus and WHOTS, while only 0.6°C warmer at NTAS, reflecting reduced downstream atmospheric advective cooling there and a smaller sensible heat flux (Table 5).

The three sites have similar mean tradewinds, westward/equatorward at 5.5 to 6.3 m s-1. The steadiest winds are at Stratus, with greatest directional variability at WHOTS. Direction deviations at WHOTS are often coincident with cool excursions in Ta, drier air, greater rainfall, and reduction in DSWR associated with midlatitude storm frontal passages in the late fall through spring, overlapping a late summer-fall tropical regime with convective clouds and rain. Evident in the rainfall at WHOTS was the 13.3 mm hr-1 daily-average associated with Hurricane Darby in July 2016. NTAS has a similar rainy period from late summer through the winter. Cool air advective events are not as evident in the one-day averages there nor is there as much directional variability in the wind, but there are drops in DSWR coincident with rain point to deep convective cloud systems. The mean rain rate at NTAS is close to that at WHOTS, both an order of magnitude larger than the mean at Stratus. Stratus rarely sees convective weather events that characterize the other two ORS sites (Weller, 2015).

Figure 4 shows the daily-averaged air-sea fluxes at Stratus (wind stress, Qnet and its components, and net freshwater flux represented by cumulative precipitation and evaporation). The complementary figures for WHOTS and NTAS are in the Supplementary Material. Table 5 presents the mean long-term fluxes. All three ORSs have mean net ocean heating of between 20 and 40 W m-2. Shortwave radiation penetrates the ocean mixed layer (OML); over the record lengths OML heat advection and entrainment cooling must approximately balance the heat flux through the surface as long-term rise in SST has not been seen. Mean QLW is between -42 and -57 W m-2, largest at WHOTS corresponding to the minimum Qnet there. The QLW heat loss is larger than the turbulent sensible heat flux (-8 to -4 W m-2). Latent heat flux (QL) is the largest component of heat loss, -106.9 W m-2, -135.9 W -2and -139.0 W m-2, respectively at Stratus, WHOTS and NTAS, reflecting the downstream increases of wind speed and SST. Over 1.2 m yr-1 of freshwater is transferred to the atmosphere at each ORS site, which is returned to the ocean along the ITCZs. Although mean Qnet is into the ocean, the annual cycles have peak ocean heating in summer and peak ocean cooling in winter. As net SWR (QSW) decreases into winter, Qnet swings from positive to negative. Short-lived events of greater ocean cooling in the winter months at WHOTS and NTAS are associated with cloud-related decreases in net SWR and related increases in wind speed and Qe. Mechanical forcing of the ocean surface by the mean wind stress is similar across the three sites, although strongest at NTAS and most variable in direction at WHOTS.

**5. Low-frequency variability and comparisons of the ORS time series to modern reanalyses and models**

A particular incentive for sustaining the ORS is to document and explore low-frequency variability (periods of a year and longer). Using the first ten years of observations Weller (2015) found significant trends in Stratus WSPD, wind stress magnitude ||, and Qe; annual mean || increased by 29% from 2000 to 2010 while annual mean Qnet decreased 39 W m-2 moving the positive annual mean closer to zero. Weller (2015) used NCEP reanalysis surface meteorological fields to relate the increasing trend in the tradewinds to a translation to the northeast and strengthening of the South Pacific high. However, the updated Stratus time series show that the increasing wind stress trend ended in 2010 and was followed by one to three-year periods of both decreasing and increasing winds. The Stratus net annual heating of the ocean returned to about 40 W m-2 over the remainder of the record through 2020. The lesson here seems obvious ― linear climate trend estimates from short records can be misleading especially in regions with large quasi-decadal natural variability.

Figure 5a includes low-passed (365-day running mean) wind stress magnitude for Stratus, WHOTS, and NTAS; Figure 5b shows the low-passed net air-sea heat flux. WHOTS low-passed || was close to 0.09 N m-2 through 2010, showed an increase in late 2011 to early 2012 then fell through early 2013 before increasing back to close to 0.09 N m-2 in 2016 and later. The WHOTS ORS showed an increasing low-passed Qnet from 2004 to 2011 followed by a period of diminishing ocean heating from 2011 to 2018. In contrast the low-passed || at NTAS showed a different temporal variability; an initial decrease from 2002 to 2005 was followed by monthly values close to the mean through late 2016, with a brief dip in 2010. As with ||, the NTAS low-passed Qnet was relatively constant over the record. The variability seen at the Pacific ORS points to a quasi-decadal cycle (cf. Meehl et al., 2008), confounding trend estimates, while the variations at NTAS seem dominated by ENSO (Newman et al., 2016).

Gridded reanalysis fields establish the spatial context for the ORS' temporal variability. In addition, another possible use would be to fill gaps in the ORS time series. With these in mind, || and Qnet heat from the ERA5, MERRA2, and NCEP2 reanalyses were compared to ORS observations. Surface meteorology and air-sea fluxes were extracted from the ERA5 (Hersbach et al. 2020), MERRA2 (Gelaro et al 2017), and NCEP2 (Kanamitsu et al, 2002; Saha et al. 2014) reanalyses at grid points near each ORS. In parallel mean SSTs from the historical runs of 51 models of the sixth Coupled Model Intercomparsion Project (CMIP6, Eyring et al., 2016) were obtained. Fifteen of those models also provided Qnet, allowing comparison with ORS heat flux.

Figure 5a compares 365-day running means of the daily wind stress from the three ORS with that from the three reanalyses, truncated to avoid filter end effects. Much of the low-passed temporal variability of │τ│ is common to the reanalyses and the ORS records, but there are distinct biases, and NCEP2 differs from the other two. MERRA2 ||is the closest of the three reanalyses to the three ORS time-series. ERA5 and MERRA2 are consistently higher than observed. Record-length means for the ORS and three reanalyses are also shown in Figure 5a. NCEP2 || at WHOTS stands out for a transition in 2014-2015 that moves from a strong low bias to align more closely with the WHOTS ORS even though that data is withheld from the analysis product. At Stratus the low bias in NCEP2 || is also reduced midway through the record in 2011 but returns in 2016. At the same time, the high biases in ERA5 and NCEP2 at Stratus are reduced, and these reanalyses track the Stratus ORS || going forward.

Figure 5b compares low-passed Qnet at the three ORS with the three reanalyses and provides the record-length means. ERA5 mean Qnet is biased low at all three ORSs. At NTAS, the three reanalyses' Qnet are closer to each other than to the ORS observations and all are biased low by close to 40 W m-2. At WHOTS, NCEP2 and MERRA2 have periods during which they track the ORS Qnet but also periods of low bias. At Stratus NCEP2 shows the initial decade of decreasing low-passed Qnet but overestimates Qnet from 2005 to 2011, after which it joins the other two reanalyses with a low bias of about 30 W m-2. At all ORS there are times when the low-passed reanalyses show oceanic heat loss while the ORS observations always show ocean heating. ERA5 mean Qnet at WHOTS stands out with a mean ocean heat loss of -4.4 W m-2 compared to the observed mean gain of 25.3 W m-2; across the ORS sites reanalysis mean Qnet values range from 27% to 117% low compared to the ORS means.

Coincident time periods of ORS data and CMIP6 results were selected and the mean SST at each ORS was compared with each of the 51 model runs as well as the ensemble mean CMIP6 SST at each site. Three time periods were chosen: for Stratus (10/18/2000 to 4/8/2018), for WHOTS (8/13/2004 to 9/15/2018), and for NTAS (3/31/2001 to 6/12/2018). Where net air-sea heat flux was available (unfortunately only from 15 models) the means for each model and the ensemble model mean at each ORS were computed. Figure 7 summarizes the mean SSTs from the historical simulations in the upper panel, one filled circle for each model, color coded for each ORS. The dashed lines are the observed mean SSTs: 20.44°C for Stratus, 25.08°C for WHOTS and 27.18°C for NTAS. The stars are the means of the 51 CMIP-6 models: 21.16°C for Stratus, 24.56°C, for WHOTS, and 26.53°C for NTAS. The model ensemble means are 0.82°C warmer at Stratus, 0.52°C cooler at WHOTS and 0.65°C cooler at NTAS. Looking across the individual model mean SSTs, model SST biases are generally consistent across the models at each site ― too cool at WHOTS and NTAS, too warm at Stratus, but there is a spread of several degrees across the models at each ORS. Extreme individual model biases were 3.26°C, -3.02°C, and -2.62°C at Stratus, WHOTS and NTAS, respectively.

The CMIP6 models' ensemble mean net-air sea heat flux is biased low at all three sites, from 18 to 55 Wm-2 too low. The 15-model mean Qnet (and all but one model) shows ocean cooling at WHOTS. The 15-model mean Qnet is near zero at NTAS. The 15-model mean Qnet shows small ocean heating at Stratus. The biases (model - ORS) between the 15-model mean and the ORS Qnet are -30.4 (Stratus), -40.0 (WHOTS) , and -36.1 Wm-2 (NTAS). Although there is a scatter of ~20 W m-2 in the individual model mean Qnet values at each site, no model mean comes closer than about 15 W m-2 to the observed ORS means. This subset of 15 models also shows ensemble mean SSTs that are too warm at Stratus and too cool at WHOTS and NTAS. The consistent underestimation of air-sea heat flux at all three sites while two sites have a cold SST bias and one site a warm SST bias suggests a regionality to the performance of the CMIP6 models and a greater complexity to processes that govern SST than surface heat flux alone. There is a suggestion in the scatter of the circles for Stratus and NTAS that warmer mean SST tends to be associated with greater low bias in Qnet.

**6. Discussion and Conclusions**

The extra-equatorial tradewind regime was targeted by the ORS effort for its important role in weather and climate and sparse measurement coverage. Easterly trade winds cover roughly 50% of the ocean surface, in a wide belt spanning the equator and extending into the subtropics. Various studies of reanalysis fields and with coupled climate models have highlighted long-term increases in the width, but not strength, of the Hadley Cell (Staten et al., 2020) and with mixed results regarding the strength of the near-equatorial Walker Circulation component of the tradewinds in the Pacific (Zhao and Allen, 2019). The data from the three ORS in the tradewinds are presented here to make their availability and the supporting assessments of the accuracies of the *in-situ* observations in the tradewinds known. At the same time, comparisons with reanalyses and coupled climate models are done to point to the need to understand the realism of these models.

Significant differences were found between observed || and Qnet and the ERA5, MERRA2, and NCEP2 reanalyses. Some differences may stem from changes in the models, while others may arise from time-varying density and types of assimilated observations. NCEP, for example, has made improvements to their Climate Forecasts System Reanalysis (CFSR) which may explain changes in NCEP2 noted in 2011 and 2014 in Figures 5a and 5b (Saha et al, 2014). In an ongoing work we look at the sources of differences, which can stem from both the algorithms used to compute the fluxes and the basic variables. Additionally, the reliance on bulk formulae for estimating fluxes continues to be addressed (Cronin et al., 2019). Observing packages with fast response sensors that enable direct computation of the turbulent fluxes are now being deployed on some buoys in parallel with the ASIMET systems and are yielding updates of the COARE flux algorithms (Edson et al., 2013). The ORS fluxes presented here were computed using COARE 3.0. Versions 3.5 and 3.6 are available (ftp1.esrl.noaa.gov and https://github.com/noaa-psd/pyCOARE). The ORS fluxes were computed again using COARE 3.6; the mean difference from COARE 3.0 for the ORS was 0.13 W m-2 for QL and -0.08 W m-2 for sensible heat flux (QS). COARE 3.6 yielded mean |t| that was 4% lower. The updates to the COARE algorithm do not significantly impact the findings.

The ORS time series allow identification of significant biases found in reanalyses and in coupled models. Reanalysis products have some temporal changes not seen in withheld observations. Care is recommended in using gridded reanalysis surface fields over the ocean. Significant biases in SST and Qnet were found in the historical run of the CMIP6 models, also suggesting care should be taken in using these models to examine the surface energy balance over the ocean and the evolution of SST. It is important to understand which of the heat flux components contribute most to the differences and consider the possible cancelation of errors among the four components. Mean Qnet values in Fig 6a are small in comparison to the magnitudes of mean Qe and QSW, comparable to those of mean QLW, and larger than those of Qs. Thus, Qnet has a particular sensitivity to accurate determination of Qe and QSW. Summing the differences of Qnet and each heat flux component across the three sites and three reanalyses, the overall reanalysis tradewind Qnet bias ranges from ~-10 to -30 W m-2 (too little into the ocean) which is characteristic of uncoupled atmospheric models when SST is specified (cf. Bretherton and Battisti, 2000). This bias is often a result of up to ~30 Wm-2 excess Qe extracted from the ocean in the reanalyses, QLW lost from the ocean that is ~5 to 15 Wm-2 too large in the reanalyses, and QSW into the ocean being up to ~15 Wm-2 too large in the reanalyses, accompanied by small sensible heat losses from the ocean that are too large in the reanalyses. While there are generally consistent differences (Fig. 6b) between reanalysis values and ORS observations across the three sites (e.g. Qe), there are a few distinct outliers within one product at a particular site (e.g. NCEP2 QSW too small at WHOTS).

The tradewind ORS-observed ocean surface energy fluxes reveal important differences of reanalyses and coupled model simulations from reality. Such consistent differences must then accumulate over thousands of km within the ocean flowing westward under the tradewinds. Mixed layers (ML) that are too deep would moderate the accumulated errors along ocean ML streamlines. Given the models’ underestimation of net heat flux into the ocean at each site, the accumulated heat budget errors could only be offset in coupled models by underestimating turbulent entrainment or underestimating advective cooling.

While we compared the results of 51 CMIP-6 ensemble-member model runs with the ORS-observed SST, and the 15 historical runs that reported Qnet, it is apparent that coupled models with large SST biases do not add anything to the simulation of air-sea interaction in the tradewinds. The use of ORS air-sea fluxes can help identify, and focus future efforts on, the most realistic models and reanalyses. We highlight Qnet as an essential metric that should be a required variable in all future CMIP efforts.

The oceanographic community is working together (https://airseaobs.org) to increase sampling of surface meteorology and air-sea interactions over the ocean during the UN Decade of the Ocean Science for Sustainable Development Program (https://www.oceandecade.org). As surface meteorological and air-sea flux data become more available, synergistic efforts among the observing, modeling, and remote sensing communities should improve utilization of observations to improve models. We hope the descriptions here of the ORS observing methods, of the production of quality-controlled time series of surface meteorology and air-sea fluxes, and of the quantification of associated uncertainties motivate other investigators to use the ORS data. In addition to the high quality of the ORS observations of air-sea interaction in the tradewind regions presented here, we emphasize that the sites were chosen to be representative of those regions. For the annual-average and low-pass-filtered observations, the results are broadly applicable to large areas. The discrepancies of reanalyses and coupled model runs from ORS averages cannot be attributed to grid-scale gradients and weather.

**Acknowledgements**

The Ocean Reference Stations have been supported by the National Oceanic and Atmospheric Administration (NOAA) from their inception. Present support for the ORS and RAW, AJP, and SB is from NOAA’s Global Ocean Observing and Monitoring Program under CINAR  
cooperative agreement as I understand it, NA14OAR4320158. CWF is also supported by NOAA Global Ocean Observing and Monitoring Program. The three ORS are part of the international OceanSITES program that coordinates collection of time series observations in the open ocean. The WHOTS ORS and RL and JP have been partially supported by the National Science Foundation (NSF), under the Hawaii Ocean Time-series program grants (OCE-0327513, 0926766, and OCE-126014) and by the State of Hawaii. The development of the IMET/ASIMET system reflects the talents of Ken Prada, Geoff Allsup, Dave Hosom and others. The maintenance of the three ORS depends critically on the skill of the technical staff, now Emerson Hasbrouck, Ray Graham, and others at WHOI and F. Santiago-Mandujano and J. Snyder, now Dan Fitzgerald at the University of Hawaii. The WHOI meteorological sensor calibration facility and sensor testing are done by Jason Smith. Data quality control and analysis support by Kelan Huang has been essential. The use of the ERA5 reanalysis from the European Centre for Medium Range Weather Forecasts is gratefully acknowledged as is the serving of that data by the Copernicus Programme.

**Data Availability Statement**

The three ORSs are part of OceanSITES, a component of the Global Ocean Observing System. The data are submitted to the OceanSITES Global Data Assembly Center at the NOAA National Data Buoy Center (<http://dods.ndbc.noaa.gov/oceansites>), which is mirrored by the Coriolis system at Ifremer (ftp://ftp.ifremer.fr/ifremer/oceansites/. The ORS data are available as merged time series at http://uop.whoi.edu/ReferenceDataSets/index.html; each of the three ORS also have their own pages on the uop.whoi.edu website. ERA5 hourly data on pressure levels from 1979 to present were downloaded from Copernicus:

<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form>

NCEP2 6-hourly data (from dataset ds094.0), NCEP Climate Forecast System Version 2 (CFSv2) 6-hourly Products, were downloaded from UCAR: https://rda.ucar.edu/datasets/ds094.0/. NASA MERRA2 hourly data are available at: <https://goldsmr4.gesdisc.eosdis.nasa.gov/opendap/MERRA2/M2T1NXOCN.5.12.4/>

<https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/M2T1NXOCN.5.12.4> and

<https://disc.gsfc.nasa.gov/datasets/M2I6NVANA_5.12.4/summary>.

CMIP6 output was accessed via the NCAR ESM using scripts modified from those provided by the PANGEO project (https://raw.githubusercontent.com/NCAR/intake-esm-

**References**

Albrecht, B.A., 1979: A model of the thermodynamic structure of the trade-wind boundary layer: Part II. Applications*. J. Atmos. Sci.,* **36(1),** 90-98, https://do.org/10.1175/1520-0469(1979)036

Albrecht, B.A., 1984: A model study of downstream variations of the thermodynamic structure of the trade winds. *Tellus*, **36A**, 187-202, [**https://doi.org/10.1111/j.1600-0870.1984.tb00238.x**](https://doi.org/10.1111/j.1600-0870.1984.tb00238.x)

Albrecht, B. A., C. S. Bretherton, D. Johnson, W. H. Schubert and A. S.Frisch, 1995a: The Atlantic Stratocumulus Transition Experiment—ASTEX. *Bull. Amer. Meteor. Soc.,* **76(6),** 889-904. https://journals.ametsoc.org/view/journals/bams/76/6/1520-0477\_1995\_076\_0889\_taste\_2\_0\_co\_2.xml

Albrecht, B.A., M.P. Jensen and W.J. Syrett, 1995b: Marine boundary layer structure and fractional cloudiness. *J. Geophys. Res.,* **100**, D7, 14,209-14,222.

Albrecht, B., Ghate, V., Mohrmann, J., Wood, R., Zuidema, P., Bretherton, C., Schwartz, C., Eloranta, E., Glienke, S., Donaher, S., Sarkar, M., McGibbon, J., Nugent, A. D., Shaw, R. A., Fugal, J., Minnis, P., Paliknoda, R., Lussier, L., Jensen, J., Vivekanandan, J., Ellis, S., Tsai, P., Rilling, R., Haggerty, J., Campos, T., Stell, M., Reeves, M., Beaton, S., Allison, J., Stossmeister, G., Hall, S., & Schmidt, S. , 2019: Cloud System Evolution in the Trades (CSET): Following the Evolution of Boundary Layer Cloud Systems with the NSF–NCAR GV. *Bull. Amer. Meteor. Soc*., **100(1),** 93-121. https://journals.ametsoc.org/view/journals/bams/100/1/bams-d-17-0180.1.xml

Bourlès, B., M. Araujo, M. J. McPhaden, P. Brandt, G. R. Foltz, R. Lumpkin, H. Giordani, F. Hernandez, N. Lefevre, P. Nobre, E. Campos, R. Saravanan, J. Trotte-Duha, M. Dengler, J. Hahn, R. Hummels, J. F. Lubbecke, M. Rouault, L. Cotrim, A. Sutton, M. Jochum, and R. C. Perez, 2019: PIRATA: A sustained observing system for tropical Atlantic climate research and forecasting. *Earth and Space Sci.*, **6**, 577-616, https://doi.org/10.1029/2018EA000428.

Bretherton, C.S., S.K. Krueger, M.C. Wyant, P. Bechtold, E. Van Meijgaard, B. Stevens and J. Teixeira, 1999: A GCSS Boundary-Layer Cloud Model Intercomparison Study of The First ASTEX Lagrangian Experiment. *Bound.-Lay. Met.,* **93,** 341–380. https://doi.org /10.1023/A:1002005429969

Bretherton, C. S., and D. S. Battisti, 2000: An interpretation of the results from atmospheric general circulation models forced by the time history of the observed sea surface temperature distribution. *Geophys. Res. Lett*., **27**, 767–770, https://doi.org/:10.1029/1999GL010910.

Colbo, K. and R.A. Weller, 2000: Accuracy of the IMET sensor package in the subtropics. *J. Atmos. Oceanic Technol.,* **26**, 1867-1890. DOI: 10.1175/2009JTECHO667.1

Cronin, M.F., C.L. Gentemann, J. Edson, I. Ueki, K. Ando, M. Bourassa, S. Brown, C.A. Clayson, C.F. Fairall, T. Farrar, S.T. Gille, S. Gulev, S.A. Josey, S. Kato, M. Katsumata, E. Kent, M. Krug, P.J. Minnett, R. Parfitt, R.T. Pinker, P.W. Stackhouse Jr., S. Swart, H. Tomita, D. Vandemark, R.A. Weller, K. Yoneyama, L. Yu, and D. Zhang. Air-sea fluxes with a focus on heat and momentum. OceanObs'19, *Front. Mar. Sci.,* doi: 10.3389/fmars.2019.00430, 2019.

Davis, R. E., L. D. Talley, D. Roemmich, W. B. Owens, D. L. Rudnick, J. Toole, R. Weller, M.J. McPhaden and J. A. Barth, 2018: 100 Years of Progress in Ocean Observing Systems, *Meteor. Monogr.*, **59**, Amer. Meteor. Soc., 3.1–3.46.

Edson, J., V. Jampana, R. Weller, S. Bigorre, A. Plueddemann, C. Fairall, S. Miller, L. Mahrt, D. Vickers, and H. Hersbach, 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>

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E., 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.,* **9,** 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016..

Fairall, C. W., E. F. Bradley, J. S. Godfrey, G. A. Wick, J. B. Edson, and G. S. Young, 1996: Cool-skin and warm-layer effects on sea surface temperature. *J. Geophys. Res.,* **101**, 1295–1308, https://

doi.org/10.1029/95JC03190.

——, ——, 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.

Fairall, C. W., T. Uttal, D. Hazen, J. Hare, M. F. Cronin, N. Bond and D. E. Veron, 2008: Observations of Cloud, Radiation, and Surface Forcing in the Equatorial Eastern Pacific, *J. Climate.*, **21(4),** 655-673.

**Gelaro, R. et al, 2017: MERRA-2 Overview:**[The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)](https://doi.org/10.1175/JCLI-D-16-0758.1), *J. Climate*.,[doi: 10.1175/JCLI-D-16-0758.1](https://doi.org/10.1175/JCLI-D-16-0758.1" \t "_blank)

Hersbach, H. *et al*., 2020: The ERA5 global reanalysis. *Q J R Meteorol Soc.,* **146**, 1999-2049, <https://doi.org/10.1002/qj.3803>.

Hosom, D.S., R.A. Weller, R.E. Payne, and K.E. Prada, 1995: The IMET (improved meteorology) ship and buoy systems. *J. Atmos. Oceanic Technol*., **12**, 527–540, [https://doi.org/10.1175/1520-0426(1995)012<0527:TIMSAB>2.0.CO;2](https://doi.org/10.1175/1520-0426(1995)012%3c0527:TIMSAB%3e2.0.CO;2)

Kanamitsu, Masao, Wesley Ebisuzaki, Jack Woollen, Shi-Keng Yang, J. J. Hnilo, M. Fiorino, G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643. doi:[10.1175/BAMS-83-11-1631](http://dx.doi.org/10.1175/BAMS-83-11-1631)

McPhaden, M. J., 1995: The Tropical Atmosphere-Ocean array is completed. *Bull. Amer. Meteor. Soc.*, **76**, 739–741, <https://doi.org/10.1175/1520-0477-76.5.739>.

McPhaden, M.J., G. Meyers, K. Ando,Y. Masumoto, V.S.N. Murty, M. Ravichandran, F. Syamsudin, J. Vialard, L. Yu, and W. Yu, 2009: RAMA: The Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction. *Bull. Am. Meteor. Soc.,* ***90****,* 459-480, <https://doi.org/10.1175/2008BAMS2608.1>

Meehl, G.A., J. M. Arblaster, G. Branstator, and H. van Loon, 200: A coupled air-sea response mechanism to solar forcing in the Pacific region. *J. Climate,* **21**, 2883-2897.

Nespor, V. and B. Sevruk, 1999: Estimation of wind-induced error of rain gauge measurements using a numerical simulation. *J. Atmos. Oceanic Technol.,* **36**, 450-464.

Newman, M. M. A. Alxeander, T. R. Ault, K.M. Cobb, C. deser, E. DiLorenzo, N. J. Mntua, A. J. Miller, S. Minobe, H. Nakamura, N. Schneider, D. J. Vimont, A.S. Phillips, J. D. Scot, and C.A. Smith, 2016: The Pacific decadal oscillation, revisited. *J. Climate*, **29**, 4399-4427.

Pinker, R. T., B. Zhang, R. A. Weller, and W. Chen, 2018: Evaluating surface radiation fluxes observed from satellites in the Southeastern Pacific Ocean. *Geophys. Res. Let.*, **45**, 2404-2412, <https://doi.org/10.1002/2017GL076805>

Saha, S. and co-authors, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185-2208. DOI: 10.1175/JCLI-D-12-00823.1

Staten, P. W., Grise, K. M., Davis, S. M., Karnauskas, K., & Davis, N., 2019: Regional widening of tropical overturning: Forced change, natural variability, and recent trends. *J. Geophys. Res. Atm.*, **124**, 6104– 6119. https://doi.org/10.1029/2018JD030100

Staten, P. W., Grise, K. M., Davis, S. M., Karnauskas, K. B., Waugh, D. W., Maycock, A. C., Fu, Q., Cook, K., Adam, O., Simpson, I. R., Allen, R. J., Rosenlof, K., Chen, G., Ummenhofer, C. C., Quan, X., Kossin, J. P., Davis, N. A., & Son, S., 2020: Tropical Widening: From Global Variations to Regional Impacts, *Bull. Amer. Meteor. Soc.,* **101**(6), E897-E904. https://journals.ametsoc.org/view/journals/bams/101/6/BAMS-D-19-0047.1.xml

Sutton, A.J., C.L. Sabine, S. Maenner-Jones, N. Lawrence-Slavas, C. Meinig, R.A. Feeley, J. T. Mathis, S. Musielewicz, R. Bott, P.D. McLain, H.J. Fought, and A. Kozyr, 2014: A high-frequency atmospheric and seawater *p*CO2 data set from 14 open-ocean sites using a moored autonomous system. *Earth Syst. Sc. Data.*, **6**, 353-366. https://doi.org/10.5194/essd-6-353-2014

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

Weller, R.A., S. P. Bigorre, J. Lord, and J. D. Ware, 2012: A surface mooring for air-sea interaction research in the Gulf Stream. Part I: Mooring design and instrumentation. *J. Atmos. Oceanic Technol.*, **29**, 1363-1376, https://doi.org/10.1175/JTECH-D-12-00060.1

Weller, R.A., 2015: Variability and Trends in Surface Meteorology and Air–Sea Fluxes at a Site off Northern Chile. *J. Climate*, **28**, 3004–3023' http://dx.doi.org/10.1175/JCLID-14-00591.1

Weller, R. A., 2016: Observing surface meteorology and air-sea fluxes. *Observing the Oceans in Real Time – Instruments, Measurements and Experience.* R. Venkatsen et al.,Eds.,Springer, 17-35. DOI 10.1007/978-3-319-66493-4\_2.

Yu, L. and R. A. Weller, 2007: Objectively Analyzed air-sea heat Fluxes (OAFlux) for the global ice-free oceans. *Bull. Amer. Meteor. Soc.*, **88**(4), 527-540, <https://doi.org/10.1175/BAMS-88-4-527>.

Zhao, X. and R. J. Allen,2019: Strengthening of the Walker Circulation in recent decades and the role of natural sea surface temperature variability. *Environ. Res. Commun.,* **1,** https://doi.org/10.1088/2515-7620/ab0dab

Zuidema, P. and co-authors (The U.S. CLIVAR Eastern Tropical Oceans Synthesis Working Group), 2016: Challenges and prospects for reducing coupled climate model SST biases in the eastern tropical Atlantic and Pacific Oceans. *Bull. Amer. Meteor. Soc*., **97**, 2305-2307. https://doi.org/:10.1175/BAMS-D-15-00274.1

**Tables**

|  |  |  |  |
| --- | --- | --- | --- |
| **Observable** | **Sensor make and model** | **Typical height above sea surface** | **Notes** |
| Wind | RM Young 5103 | 2.82 m | Propeller-vane anemometer, stock propeller bearing upgraded |
| Wind | Gill Instruments WindObserver II Ultrasonic Anemometer | 3.08 m | Sonic anemometer, used on at times to mitigate data loss due to birds |
| Air temperature/humidity | Rotronic MP-101A | 2.95 m | Porous Teflon filter and multiplate radiation shield |
| Incoming shortwave radiation | Eppley Precision Spectral Pyranometer | 3.43 m | Case adapted to ASIMET module tubing |
| Incoming longwave radiation | Eppley Precision Infrared Radiometer | 3.43 m | Case adapted to ASIMET module tubing |
| Barometric pressure | Heise DXD | 3.0 m | With parallel plate pressure port |
| Precipitation | RM Young 50202 | 3.12 m | Self-siphoning rain gauge |
| Sea surface temperature and salinity | SeaBird 37 | -0.85m | Mounted on buoy bridle |

Table 1. Present sensors in the ASIMET system. Each are configured as an ASIMET module, capable of internal recoding and also wired to a central logger.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sensor** | **Instantaneous** | **Daily** | **Annual** |
| Downward longwave (W m-2) | 7.5 | 4 | 4 |
| Downward shortwave (W m-2) | 20 | 6 | 5 |
| Humidity (%RH) | 1  3 (low winds) | 1  3 | 1 |
| Air temperature (°C) | 0.2 (more in low wind) | 0.1 | 0.1 |
| Barometric pressure (hPa) | 0.3 | 0.2 | 0.2 |
| SST (°C) | 0.1 | 0.1 | 0.004 |
| Wind speed (m s-1) | 1.5% or 0.1  (more in low wind) | 1%, 0.1  (max of these) | 1%, 0.1  (max of these) |
| Wind direction (°) | 6 (more in low wind) | 5 | 5 |
| Rainfall (% under catchment) | 10 | 10 | 10 |

Table 2. Summary of ASIMET sensor accuracies in the field after Colbo and Weller (2000) and Weller (2016). The 20 W m-2 value in instantaneous downward shortwave reflects finding shortwave differences in the one-minute sampled data. Short-lived transient differences are observed in adjacent DWSR modules as small clouds pass over; the differences stem primarily from time base (clock) differences rather than pyranometer performance.

|  |  |  |
| --- | --- | --- |
| **Flux** | **Typical** | **Percent Error** |
| Net longwave (W m-2) | 3.9 | 10 |
| Net shortwave (W m-2) | 5 | 2.5 |
| Latent (W m-2) | 5 | 5 |
| Sensible (W m-2) | 1.5 | 15 |
| Net heat flux (W m-2) | 8 | 20 |
| Wind stress magnitude (N m-2) | 0.007 | 20 |
| *E-P* (cm) | 10 | 10 |

Table 3. Accuracies of annual averaged fluxes from an ORS buoy. After Colbo and Weller (2000)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Stratus** | **WHOTS** | **NTAS** |
| Wind speed (m s-1) | 5.8 | 5.5 | 6.3 |
| Wind dir (towards, °) | 304.8 | 263.2 | 257.4 |
| Air temp (°C) | 19.6 | 24.2 | 26.6 |
| SST (°C) | 20.5 | 25.1 | 27.2 |
| RH/SH (%/g kg-1) | 73.4/10.4 | 75.5/14.2 | 76.6/16.6 |
| Barometric press (hPa) | 1017.7 | 1017.0 | 1014.6 |
| Downward shortwave (W m-2) | 205.2 | 237.3 | 246.7 |
| Downward longwave (W m-2) | 376.8 | 388.8 | 405.9 |
| Rain rate (mm hr-1) | 0.007 | 0.063 | 0.057 |
| Surface current speed (m s-1) | 0.04 | 0.06 | 0.08 |
| Surface current dir (°) | 263.4 | 289.9 | 300.3 |

Table 4. Long-term means of surface meteorology at the three ORS formed by taking the annual means of the mean daily time series formed from the calendar years of complete data. WHOTS has 14 full years (2005-2019), Stratus 15 full years (between 2000 and 2021), and NTAS 16 full years (between 2001 and 2021). Surface current data come from the current meter closest to the surface on the mooring and are used when computing the wind relative to the surface current in the bulk flux formulae. The variables are as observed at the heights given in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Flux** | **Stratus** | **WHOTS** | **NTAS** |
| Wind stress magnitude (N m-2) | 0.075 | 0.076 | 0.081 |
| Wind stress dir (towards, °) | 304.9 | 261.3 | 256.2 |
| Net heat flux (W m-2) | 36.9 | 24.7 | 36.2 |
| Latent heat (W m-2) | -106.9 | -135.9 | -139.0 |
| Sensible heat (W m-2) | -7.6 | -7.0 | -3.9 |
| Net shortwave (W m-2) | 193.9 | 224.3 | 231.9 |
| Net longwave (W m-2) | -42.5 | -56.5 | -52.8 |
| Freshwater (m yr-1) | -1.3 | -1.2 | -1.3 |

Table 5. Means of the long-term records of the air-sea fluxes at the three ORS formed by averaging the annual means of the mean daily time series from only the full calendar years used to make the means in Table 4. Freshwater flux is evaporation+precipitation; evaporation is negative, precipitation is positive; and heat flux and freshwater flux are negative for ocean loss.

**Figure Caption List**

Figure 1. Map showing the locations of the three Ocean Reference Stations, WHOTS, Stratus, and NTAS, plotted on top of color contours of the 2000-2020 mean sea surface temperature field and wind vectors from the ERA5 reanalysis.

Figure 2. ORS surface buoys with older, 3-m aluminum hull (left) and present 3-m foam hull (right). A wind vane with radar reflector and data telemetry antenna orients the buoy with respect to the wind, and anemometers and air temperature and humidity modules are placed on the upwind side of the tower. The cluster of four radiometers, one each for shortwave and longwave from two ASIMET systems, is placed above all other structures. Rain gauges, barometric pressure sensors are placed aft of the forward face.

Figure 3. Daily averaged surface meteorology observed at Stratus. The wind vectors (towards) are subsampled every 5 days for clarity. Gaps in the data resulted from breaks in the mooring. From the top, sea surface temperature (SST), air temperature, barometric pressure, specific humidity, rain rate, incoming longwave radiation (LWR), incoming shortwave radiation (SWR), wind speed, and vector wind are shown.

Figure 4. Daily averaged air-sea fluxes as observed at WHOTS. Gaps in the data resulted from breaks in the mooring. From the top, the magnitude of the wind stress, net air-sea heat flux (black - daily, red – 10-day running mean), latent (black) and sensible (red) heat flux, net shortwave radiation, net longwave radiation, and an overplot of cumulative rainfall and evaporation times -1. To compute sensible and latent heat flux gap-free wind speed data were used.

Figure 5a. Three panels comparing low-passed wind stress magnitude at each of the ORS (black) with that at ERA5 (blue), MERRA2 (green), and NCEP2 (red). ORS traces start/end half the width of the 365-day running mean filter from the beginning/end of data segments. Long-term means for stress in each panel are given in color-coded text.

Figure 5b. Three panels comparing low-passed net air-sea heat flux at each ORS with the three reanalyses. ORS traces start/end half the width of the 365-day running mean filter at the beginning/end of data segments. Long-term means for net heat flux in each panel are given in color-coded text.

Figure 6a. Comparisons of the long term means of net heat flux, latent heat flux, sensible heat flux, net shortwave radiation, and net longwave radiation at the three ORS (black) with the means from ERA5 (blue), NCEP2 (red), and MERRA2 (green). On the x-axis the neat heat flux and each component have a cluster of four bars labeled as W (WHOTS), S (Stratus), and N (NTAS).

Figure 6b. Comparisons of the differences (reanalysis minus ORS) in the long term means of net heat flux, latent heat flux, sensible heat flux, net shortwave radiation, and net longwave radiation at the three ORS with the means from ERA5 (blue), NCEP2 (red), and MERRA2 (green). On the x-axis the neat heat flux and each component have a cluster of three bars labeled as W (WHOTS), S (Stratus), and N (NTAS).

Figure 7. (upper) Comparison of mean SST at the three ORS with the CMIP6 models, historical simulation run; the horizontal axis is used to offset SSTs from the different CMIP6 models. The dashed lines are the ORS mean SSTs; the stars are the mean of all model SSTs. Individual models are filled circles. (lower) Comparison of both mean net air-sea heat flux and mean SST from CMIP6 models that provided both and the ORS. Squares are ORS means; stars are means off the CMIP6 models (filled circles). For both location is given by the color.

**Figures**

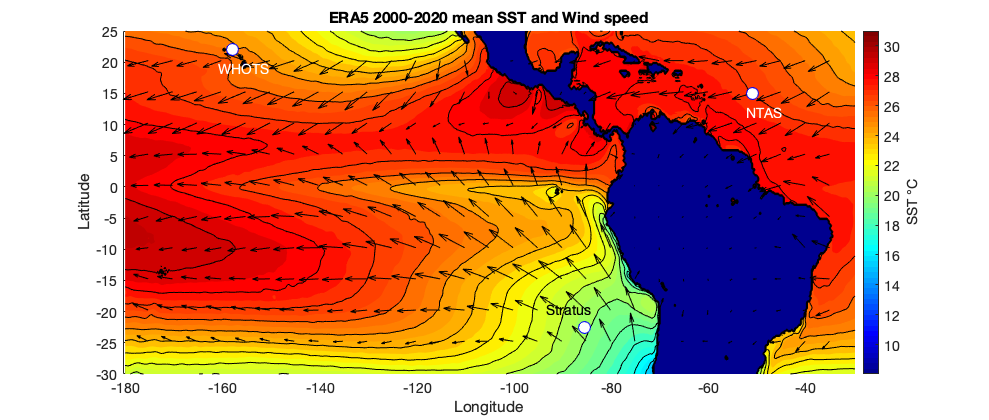
****

Figure 1. Map showing the locations of the three Ocean Reference Stations, WHOTS, Stratus, and NTAS, plotted on top of color contours of the 2000-2020 mean sea surface temperature field and wind vectors from the ERA5 reanalysis.

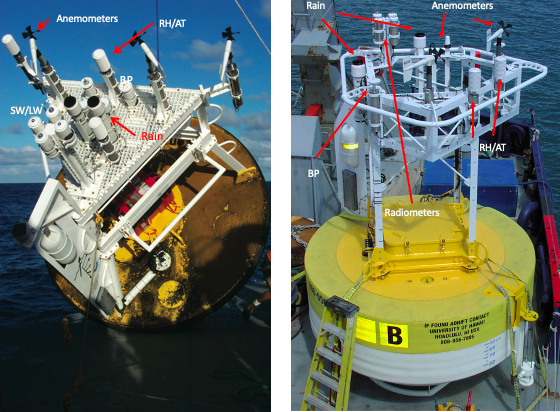


Figure 2. ORS surface buoys with older, 3-m aluminum hull (left) and present 3-m foam hull (right). A wind vane with radar reflector and data telemetry antenna orients the buoy with respect to the wind, and anemometers and air temperature and humidity modules are placed on the upwind side of the tower. The cluster of four radiometers, one each for shortwave and longwave from two ASIMET systems, is placed above all other structures. Rain gauges, barometric pressure sensors are placed aft of the forward face.

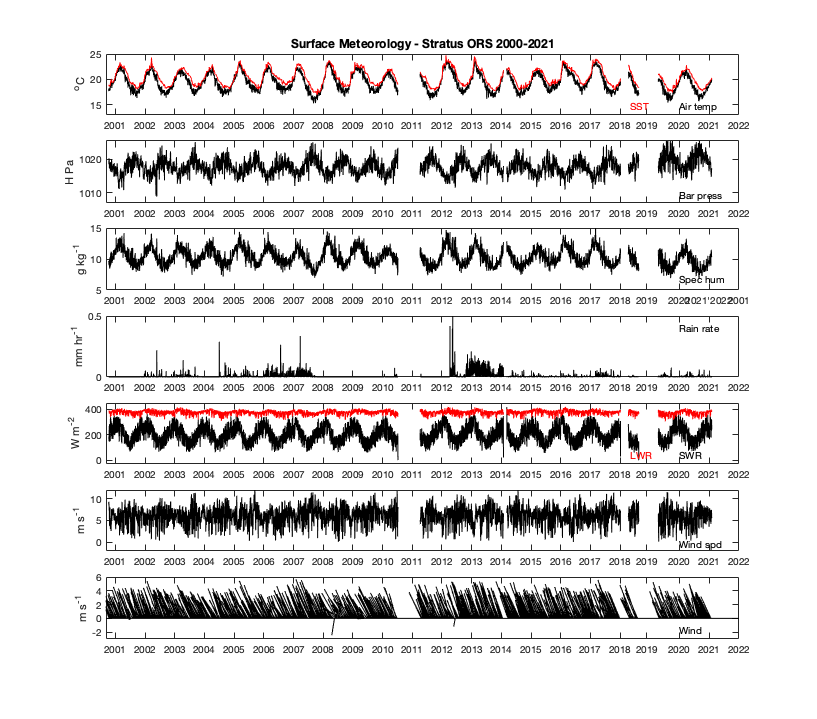


Figure 3. Daily averaged surface meteorology as observed at Stratus. The wind vectors (towards) are subsampled every 5 days. Gaps in the data resulted from breaks in the mooring. From the top, sea surface temperature (SST), air temperature, barometric pressure, specific humidity, rain rate, incoming longwave radiation (LWR), incoming shortwave radiation (SWR), wind speed, and vector wind are shown.

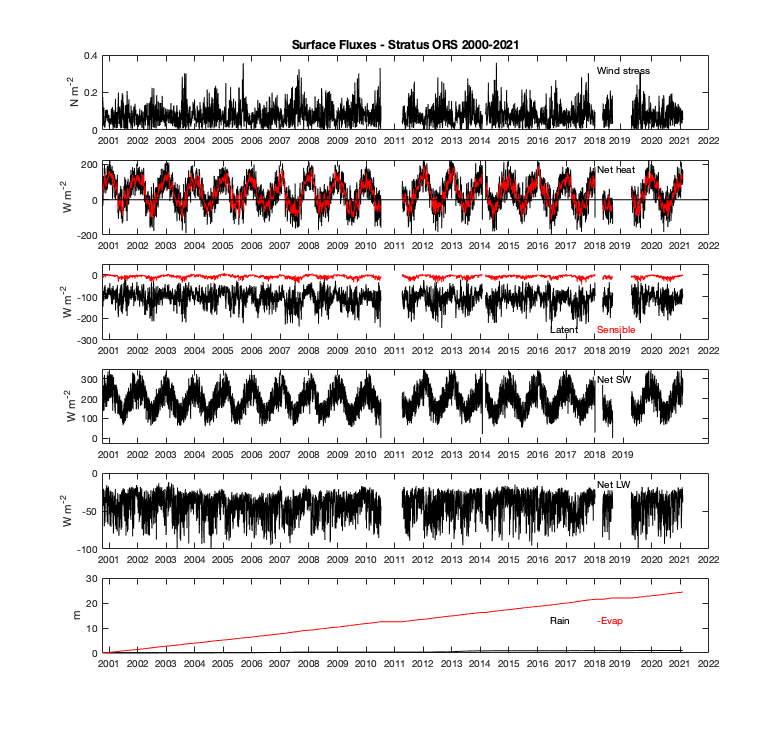


Figure 4. Daily averaged air-sea fluxes as observed at Stratus. Gaps in the data resulted from breaks in the mooring. From the top, the magnitude of the wind stress, net air-sea heat flux (black - daily, red – 10-day running mean), latent (black) and sensible (red) heat flux, net shortwave radiation, net longwave radiation, and an overplot of cumulative rainfall and evaporation.

Chart, histogram

Description automatically generated

Figure 5a. Three panels comparing low-passed wind stress magnitude at each of the ORS (black) with that at ERA5 (blue), MERRA2 (green), and NCEP2 (red). ORS traces start/end half the width of the 365-day running mean filter from the beginning/end of data segments. Long-term means for stress in each panel are given in color-coded text.

Chart, histogram

Description automatically generated

Figure 5b. Three panels comparing low-passed net air-sea heat flux at each ORS with the three reanalyses. ORS traces start/end half the width of the 365-day running mean filter at the beginning/end of data segments. Long-term means for net heat flux in each panel are given in color-coded text.

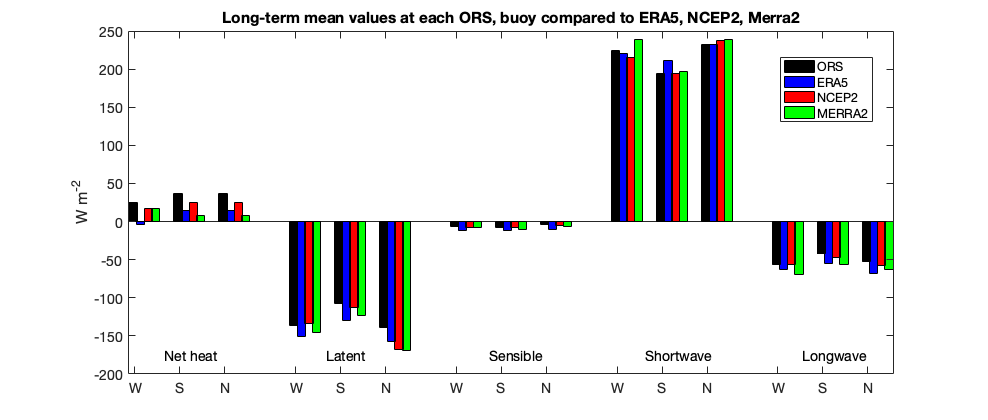


Figure 6a. Comparisons of the long term means of net heat flux, latent heat flux, sensible heat flux, net shortwave radiation, and net longwave radiation at the three ORS (black) with the means from ERA5 (blue), NCEP2 (red), and MERRA2 (green). On the x-axis the neat heat flux and each component have a cluster of four bars labeled as W (WHOTS), S (Stratus), and N (NTAS).

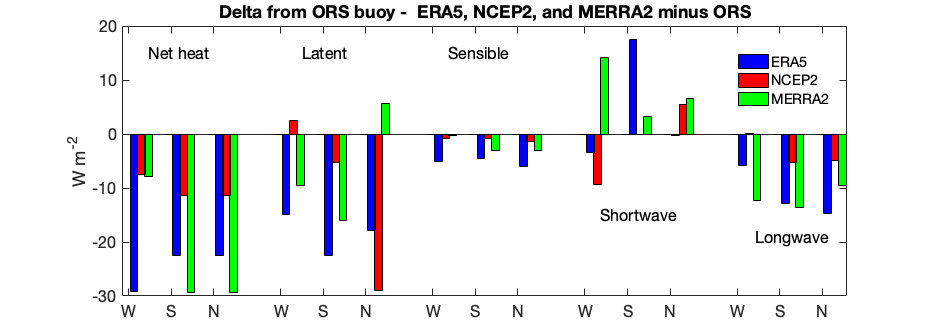
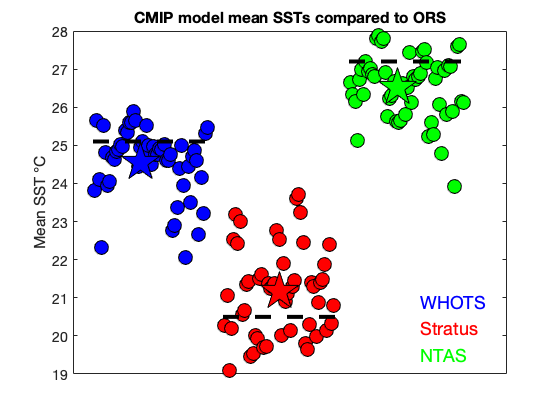


Figure 6b. Comparisons of the differences (reanalysis minus ORS) in the long term means of net heat flux, latent heat flux, sensible heat flux, net shortwave radiation, and net longwave radiation at the three ORS with the means from ERA5 (blue), NCEP2 (red), and MERRA2 (green). On the x-axis the neat heat flux and each component have a cluster of three bars labeled as W (WHOTS), S (Stratus), and N (NTAS).



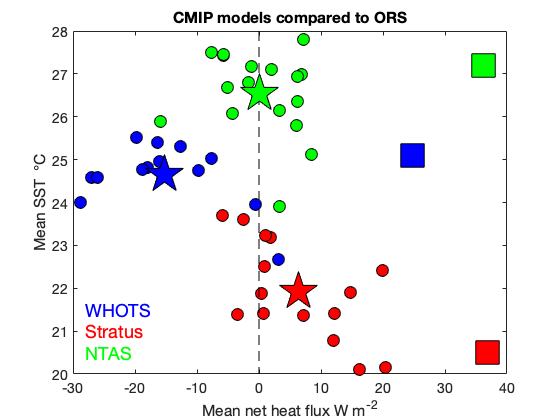


Figure 7. (upper) Comparison of mean SST at the three ORS with the CMIP6 models, historical simulation run; the horizontal axis is used to offset SSTs from the different CMIP6 models. The dashed lines are the ORS mean SSTs; the stars are the mean of all model SSTs. Individual models are filled circles. (lower) Comparison of both mean net air-sea heat flux and mean SST from CMIP6 models that provided both and the ORS. Squares are ORS means; stars are means off the CMIP6 models (filled circles). For both location is given by the color.