SPECIAL MERRA COLLECTION

An Assessment of the Uncertainties in Ocean Surface Turbulent Fluxes in 11 Reanalysis, Satellite-Derived, and Combined Global Datasets

MICHAEL A. BRUNKE, ZHUO WANG, AND XUBIN ZENG

Department of Atmospheric Sciences, The University of Arizona, Tucson, Arizona

MICHAEL BOSILOVICH

NASA Goddard Modeling and Assimilation Office, Greenbelt, Maryland

CHUNG-LIN SHIE

Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, and Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 23 November 2010, in final form 19 March 2011)

ABSTRACT

Ocean surface turbulent fluxes play an important role in the energy and water cycles of the atmosphere–ocean coupled system, and several flux products have become available in recent years. Here, turbulent fluxes from 6 widely used reanalyses, 4 satellite-derived flux products, and 2 combined product are evaluated by comparison with direct covariance latent heat (LH) and sensible heat (SH) fluxes and inertial-dissipation wind stresses measured from 12 cruises over the tropics and mid- and high latitudes. The biases range from -3.0 to 20.2 W m^{-2} for LH flux, from -1.4 to 6.0 W m^{-2} for SH flux, and from -7.6 to 7.9×10^{-3} N m^{-2} for wind stress. These biases are small for moderate wind speeds but diverge for strong wind speeds (>10 m s^{-1}). The total flux biases are then further evaluated by dividing them into uncertainties due to errors in the bulk variables and the residual uncertainty. The bulk-variable-caused uncertainty dominates many products' SH flux and wind stress biases. The biases in the bulk variables that contribute to this uncertainty can be quite high depending on the cruise and the variable. On the basis of a ranking of each product's flux, it is found that the Modern-Era Retrospective Analysis for Research and Applications (MERRA) is among the "best performing" for all three fluxes. Also, the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim) and the National Centers for Environmental Prediction–Department of Energy (NCEP–DOE) reanalysis are among the best performing for two of the three fluxes. Of the satellite-derived products, version 2b of the Goddard Satellite-Based Surface Turbulent Fluxes (GSSTF2b) is among the best performing for two of the three fluxes. Also among the best performing for only one of the fluxes are the 40-yr ERA (ERA-40) and the combined product objectively analyzed air–sea fluxes (OAFlux). Direction for the future development of ocean surface flux datasets is also suggested.

1. Introduction

The atmosphere and ocean interact at their interface through surface turbulent fluxes of temperature [sensible heat (SH)], moisture [latent heat (LH)], and momentum (wind stress τ). Knowledge of these fluxes is important to understand the ocean heat and freshwater budget and the partitioning of the global pole-to-equator heat transport between the atmosphere and the ocean. The fluxes are also needed to provide a boundary condition for both atmospheric and ocean models and are instrumental in assessing numerical weather prediction and global coupled models (Brunke et al. 2002, 2003).

Direct observation of these fluxes can be made using high-speed instruments, such as a sonic anemometer or high-speed hygrometer, but such measurements are limited to at most a few cruises per year. Another option

Corresponding author address: Michael A. Brunke, Department of Atmospheric Sciences, The University of Arizona, P.O. Box 210081, Tucson, AZ 85721-0081.
E-mail: brunke@atmo.arizona.edu

DOI: 10.1175/2011JCLI4223.1

© 2011 American Meteorological Society

is to derive fluxes from measurements of bulk quantities of air temperature, sea surface temperature, air humidity, and wind speed from moored buoys and ships of opportunities as was done for da Silva et al. (1994) and the National Oceanography Centre, Southampton (NOC), flux product (Berry and Kent 2009), which used observations from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; Worley et al. 2005). Such products utilize these bulk quantities as input into bulk aerodynamic algorithms to calculate the fluxes by way of the following equations:

$$\mathrm{SH} = \rho_a c_p C_H U(\theta_s - \theta_a), \quad (1)$$

$$\mathrm{LH} = \rho_a L_v C_E U(q_s - q_a), \quad \text{and} \quad (2)$$

$$\tau = \rho_a C_D U^2, \quad (3)$$

where ρ_a is the density of air, c_p is the specific heat of air at constant pressure, and L_v is the latent heat of vaporization. The bulk quantities are the wind speed, U; the surface and near-surface atmospheric potential temperatures, θ_s and θ_a, respectively; and the surface and near-surface atmospheric specific humidities, q_s and q_a, respectively. The values for the exchange coefficients—C_H for heat, C_E for moisture, and C_D for momentum (also known as the drag coefficient)—are determined empirically based upon the bulk quantities and differ slightly from algorithm to algorithm (e.g., see Zeng et al. 1998; Brunke et al. 2002, 2003). However, such observations of bulk quantities from moored buoys and ships of opportunity are limited temporally and spatially, being particularly absent from the Southern Hemisphere, and fluxes from these products are not very accurate (e.g., da Silva et al. 1994; Josey et al. 1999; Wang and McPhaden 2001).

A third option is to derive the bulk quantities from satellite observations, which provide truly global coverage, and the surface fluxes are generally calculated using a bulk algorithm. Several such datasets have been developed over recent years, most notably the Goddard Satellite-Based Surface Turbulent Fluxes (GSSTF; Chou et al. 2003), the Japanese Ocean Flux Data Sets with Use of Remote Sensing Observations (J-OFURO; Kubota et al. 2002), the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS, http://hoaps.org; Andersson et al. 2010), and those documented in Bentamy et al. (2003) and Bourras et al. (2002). For these datasets, wind speed and sea surface temperature can be easily derived from satellite measurements. In contrast, near-surface air humidity derived from satellite measurements is not particularly accurate, while in the past, near-surface air temperature was almost impossible to retrieve accurately (Bourras 2006). Recent studies (Jackson et al. 2006; Roberts et al. 2010; Jackson and Wick 2010), however, have produced accurate derivations of near-surface air temperature from satellite measurements. Still, some datasets use the air temperatures from reanalysis products instead of using a satellite-derived air temperature, while some others employ a constant surface–air temperature difference as done in HOAPS (Andersson et al. 2010). Another method would be to use a neural network to find a direct relationship between Special Sensor Microwave Imager (SSM/I) brightness temperatures and sensible heat flux similar to what was done for latent heat flux by Bourras et al. (2002).

Finally, sea surface turbulent fluxes can be derived from global model results that have been constrained by surface and rawinsonde observations and satellite measurements. Such products are called reanalyses and are produced by some of the major modeling centers, such as the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP), the European Centre for Medium-Range Weather Forecasts (ECMWF), the Japan Meteorological Agency (JMA), and the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) Global Modeling and Assimilation Office (GMAO). Again, bulk algorithms are utilized to calculate the fluxes generally using the values from the lowest model layer as the near-surface quantities. Because of the use of models to produce these reanalyses, model errors can be introduced into their surface fluxes (e.g., Weller and Anderson 1996; Wang and McPhaden 2001; Smith et al. 2001; Renfrew et al. 2002).

Reanalyses are widely used, while satellite-derived fluxes are becoming more widely used as their accuracy is approaching that of the reanalyses. The climatological fluxes in these products have been intercompared extensively, particularly when a new product or version has been introduced. For example, Chou et al. (2003) compared the climatological characteristics of their version 2 of GSSTF (GSSTF2) with HOAPS, the older NCEP–National Center for Atmospheric Research (NCAR) reanalysis, and da Silva et al. (1994). Similarly, Kubota et al. (2002) compared their original version of J-OFURO to da Silva et al. (1994), the NCEP–NCAR reanalysis, and ECMWF's reanalysis, and Kubota et al. (2003) added HOAPS and GSSTF to their comparison of J-OFURO2.

The goal, as established by the U.S. Climate Variability and Predictability (CLIVAR) Program and the Global Energy and Water Cycle Experiment (GEWEX), is to attain an accuracy of 5 W m^{-2} for each component of the surface heat budget (Curry et al. 2004). To get there, one must know what is contributing most to the product

bias: the algorithm used or the bulk variables inputted into them. Brunke et al. (2002) did this by separating out the algorithm and bulk variable contributions in the NASA GMAO's Goddard Earth Observing System (GEOS) reanalysis and HOAPS, and comparing them to each other. Bourras (2006) went one step further to assign contributions to the individual bulk variables for several satellite-derived products, but this was only done for latent heat flux using buoy data in the North Pacific and Atlantic.

Futhermore, with the recent availability of a new generation of several ocean surface turbulent flux products, we should now ask the question, Is the new generation of reanalyses [e.g., the Modern-Era Retrospective Analysis for Research and Applications (MERRA), the Climate Forecast System Reanalysis (CFSR), and the ECMWF interim reanalysis (ERA-Interim)] better than the previous generation [e.g., NCEP–NCAR, NCEP–Department of Energy (DOE), and 40-yr ERA (ERA-40)]? And how well do the satellite-derived and combined products perform relative to the reanalyses? Furthermore, what is the main contributor to the flux errors in all of these products? Here, we address these questions with six reanalyses, four satellite-derived products, and one product based on both satellite-derived measurements and reanalysis. These products are briefly described in section 2. Instead of using buoy measurements, as in Bourras (2006), which needed a bulk algorithm to derive the "observed" surface turbulent fluxes, we compare the product fluxes to direct observations taken from 12 experimental ship cruises in the tropics and Northern Hemisphere subtropics and mid- and high latitudes. These cruises are described briefly in section 3. To better understand the product flux biases, we split the total bias into two components, a bulk variable uncertainty and a residual one, and rank the products according to their biases and standard deviation of the errors (SDEs) as described in section 4. The results are presented in section 5, followed by some further discussion of the results and some concluding remarks in section 6.

2. Data products

The ocean surface turbulent flux data products compared here include commonly used reanalyses and satellite-derived products: NASA GMAO's MERRA, ERA-40 and ERA-Interim, NCEP's original reanalysis (versions 1 and 2) and the latest (CFSR), NASA GSFC's GSSTF versions 2 and 2b, J-OFURO version 2, HOAPS version 3, and Woods Hole Oceanographic Institution (WHOI)'s objectively analyzed air–sea fluxes (OAFlux). The bulk flux algorithms used by these products are explained in the appendix.

a. MERRA

MERRA is the most recent reanalysis produced by the NASA GMAO using the GEOS Data Assimilation System (DAS), which has at its core the GEOS version 5 (GEOS-5) atmospheric general circulation model (AGCM) (Rienecker et al. 2011; Suarez et al. 2008; Rienecker et al. 2007). The model has a finite-volume dynamical core that is run at a resolution of ½° latitude × ⅔° longitude (Suarez et al. 2008) with 72 vertical layers. Assimilation is done by gridpoint statistical interpolation (GSI), a new three-dimensional variational (3DVar) analysis (Rienecker et al. 2011). To reduce shocks from the mass-wind analysis increments, the incremental analysis update (IAU) procedure (Bloom et al. 1996) is implemented (Rienecker et al. 2007). In this procedure, assimilation at the 6-h synoptic times is based upon 6 h of model predictions centered on the synoptic time. The analyzed correction, then, is applied at the previous 6 h, and the model is run for 12 h (Suarez et al. 2008).

Conventional observational inputs to MERRA include data from surface land, ship, and buoy observations; rawinsondes; dropsondes; pilot balloons (PIBALs); wind profilers; and aircraft. Satellite inputs include upper-air winds derived from geostationary satellites and the Moderate Resolution Imaging Spectroradiometer (MODIS); surface winds from SSM/I, the Quick Scatterometer (QuikSCAT), and the European Remote Sensing Satellite (ERS) scatterometers 1 and 2 (*ERS-1* and *ERS-2*, respectively); surface rain rate from SSM/I and the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI); radiances from the Geostationary Operational Environmental Satellite (GOES) sounder, Television and Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) and Advanced TOVS (ATOVS) instruments, the Atmospheric Infrared Sounder (AIRS), the Microwave Sounding Unit (MSU), the Advanced Microwave Sounding Unit-A (AMSU-A), and SSM/I; and ozone retrievals from the Solar Backscatter Ultraviolet instrument (SBUV) (Rienecker et al. 2011). SST was taken from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST; Rayner et al. 2003) and the Reynolds et al. (2007) products.

Used here are the surface turbulent flux (tavg1_2d_flx_Nx) and single-level atmospheric bulk variable (tavg1_2d_slv_Nx) data collections, which are provided in the model horizontal resolution (½° latitude × ⅔° longitude) at 1-h temporal resolution. These data were downloaded through the MERRA Web site (http://gmao.gsfc.nasa.gov).

b. ECMWF reanalyses

Included here are two generations of ECMWF reanalyses: ERA-40 and ERA-Interim. ERA-40 utilized

the ECMWF atmospheric model with a spectral resolution of T159 and 60 vertical layers (Uppala et al. 2005). The model included improvements to the parameterizations of deep convection and radiation and a new representation of sea ice (Uppala et al. 2005). Data assimilation was done by a 3DVar system. Inputs included conventional sources (surface land and ship observations, rawinsondes, dropsondes, PIBALs, wind profilers, and aircraft) and satellite measurements [upper-level winds from geostationary satellites, radiances from the Vertical Temperature Profile Radiometer (VTPR), the High Resolution Infrared Sounder (HIRS), the Stratospheric Sounding Unit (SSU), MSU, and AMSU-A; total column water vapor and surface wind speeds from SSM/I; ocean wave height and surface wind from *ERS-1* and *ERS-2*, total column ozone from the Total Ozone Mapping Spectroradiometer (TOMS), and ozone profiles from SBUV] (Uppala et al. 2005).

ERA-Interim is the newest generation of ECMWF reanalyses (Simmons et al. 2006). Unlike ERA-40, which was limited to a 45-yr period (September 1957–August 2002; Uppala et al. 2005), ERA-Interim has near-real-time analyses. ERA-Interim's model has a higher horizontal resolution (T255) with improved model physics. For ERA-Interim, ECMWF implemented a 12-h 4DVar assimilation system with improvements to the handling of data biases and the background error constraint. The same data from ERA-40 was inputted into ERA-Interim with the addition of clear-sky radiances from *Meteosat-2*; Global Ozone Monitoring Experiment (GOME) ozone profiles; and radio occultation (RO) measurements from the Challenging Minisatellite Payload (CHAMP), the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC), and the Gravity Recovery and Climate Experiment (GRACE). Also included were reprocessed ocean wave height data from *ERS-1* and *ERS-2* and upper-level winds from *Meteosat-2* (Simmons et al. 2006).

Utilized here are the 6-hourly surface and single-level analysis on a 2.5° latitude–longitude fixed grid for ERA-40 and the model-resolution 3-hourly surface fluxes and surface bulk variables from ERA-Interim. The ERA-40 data were downloaded from the NCAR Research Data Archive (RDA) Web site (http://dss.ucar.edu), and the ERA-Interim data were downloaded from the mass storage on NCAR's supercomputers.

c. NCEP reanalyses

The NCEP reanalyses compared here include NCEP–NCAR, NCEP–DOE, and the CFSR. NCEP–NCAR (also referred to as NCEP-I) was originally a 40-yr product (Kalnay et al. 1996) but has been extended to near–real time. Its model is NCEP's operational global forecasting model of the mid-1990s with a horizontal resolution of T62 and 28 vertical layers. Assimilation is done by spectral statistical interpolation (SSI), an older 3DVar technique, using data from the conventional sources mentioned above; satellite radiances from Satellite Infrared Spectrometer (SIRS), HIRS, VTPR, and TOVS; and upper-level winds from geostationary satellites. SST was taken from the Met Office (UKMO)'s Global Sea Ice and Sea Surface Temperature dataset (GISST; Rayner et al. 2006) product before 1982 and from the Reynolds and Smith (1994) analysis afterward (Kalnay et al. 1996).

NCEP–DOE (also referred to as NCEP-2) should be considered to be a revised version of NCEP–NCAR, not a new generation of reanalysis, developed to correct errors discovered in the processing of the earlier version. As in the previous version, NCEP–DOE uses a T62 spectral resolution model with 28 layers in the vertical. Some improvements were made to the model physics, including changes to boundary layer turbulence and radiation. The ocean albedo was also reduced from 0.15 in NCEP–NCAR to 0.06–0.07 in NCEP–DOE, while desert albedo was increased in NCEP–DOE. Similar data were ingested into NCEP–DOE, except that sea ice and SST were prescribed from the Atmospheric Model Intercomparison Project (AMIP-II) analysis and a new ozone climatology (Rosenfield et al. 1987) was used. Also, a geolocation error was fixed in the Southern Hemisphere surface pressure analysis (Kanamitsu et al. 2002).

CFSR is the latest generation of NCEP reanalyses (Saha et al. 2010). It uses the Climate Forecast System (CFS), a fully coupled atmosphere–ocean–sea ice–land model. The atmospheric component is run at a spectral resolution of T382 with 64 vertical layers with the addition of a cloud microphysics scheme to determine cloud condensate prognostically (Zhao and Carr 1997; Sundqvist et al. 1989; Moorthi et al. 2001), the simplified Arakawa–Schubert cumulus convection scheme (Pan and Wu 1995; Hong and Pan 1998), and orographic gravity wave drag (Kim and Arakawa 1995; Alpert et al. 1988, 1996). The ocean component is run at a variable horizontal resolution of ¼°–½° latitude × ½° longitude and 40 layers in the vertical with the uppermost layer at 10-m thickness. Data similar to that of the other reanalyses were also ingested into the CFSR with the addition of SSM/I, ERS, QuikSCAT, and WindSat ocean surface winds; GOES, AIRS, Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E), Infrared Atmospheric Sounding Interferometer (IASI), and Microwave Humidity Sounder (MHS) radiances; and RO from CHAMP and COSMIC. SST was taken from NCEP's optimum interpolation (OI) product (Reynolds et al. 2007).

The 6-hourly surface fluxes and surface bulk variables taken from the model-resolution two-dimensional fields

of NCEP–NCAR and NCEP–DOE were downloaded from the NCAR RDA Web site, and those from CFSR's low-horizontal-resolution (~1.9° × 1.9°) hourly time series products from the mass storage on NCAR's supercomputers.

d. GSSTF products

Compared here are two of the latest versions of the NASA GSFC satellite-derived products: GSSTF2 and GSSTF2b. GSSTF2 (Chou et al. 2003) was produced for a 13.5-yr period (July 1987–December 2000) with a 1° × 1° latitude–longitude horizontal resolution and a daily temporal resolution. Wind speed is retrieved from SSM/I measurements using the version 4 (Wentz 1997) scheme. Specific humidity at 10 m is derived from SSM/I version 4 total column water vapor and from the lowest 500 m, as described in Chou et al. (1995, 1997). Near-surface air temperature and SST are taken from NCEP–NCAR. The bulk algorithm used to calculate turbulent fluxes is slightly adapted from the Chou (1993) algorithm (Chou et al. 2003; see appendix).

Recently, the GSSTF products were revived and extended with a slightly improved version— that is, GSSTF2b (Shie 2010; Shie et al. 2009) being just released using an updated version 6 of the SSM/I total precipitable water, brightness temperature, and wind speed retrieval (Wentz et al. 2007 and online supplemental material; also see online at http://www.ssmi.com) as well as NCEP–DOE near-surface air temperatures and SSTs. The horizontal and temporal resolutions are unchanged but the period has been extended out to December 2008.

Both versions of GSSTF used here were downloaded via anonymous FTP through the Goddard Earth Sciences Data and Information Services Center (http://disc.sci.gsfc.nasa.gov). GSSTF2b can also be downloaded in Hierarchical Data Format for the Earth Observing System (HDF-EOS5) format (from ftp://measures.gsfc.nasa.gov/data/s4pa/GSSTF/ or http://disc.sci.gsfc.nasa.gov/daac-bin/DataHoldingsMEASURES.pl?PROGRAM=ChungLinShie). There are two sets of data produced for GSSTF2b: sets 1 and 2 (Shie 2010). Set 1 used here was found to possess a slightly increased global latent heat flux, especially after 2000. Therefore, set 2 was produced later by removing certain available satellite products that seemed to have a relatively larger trend in latent heat flux. As such, set 1 may contain a larger global temporal trend in latent heat flux with less missing data, while there might be a smaller trend in set 2 with more missing data. Detailed information about sets 1 and 2 can be found in Shie (2010). The results here are unaffected by the choice of set 1, but there might be more of a difference if data after 1999 (when more data were disqualified) were used. The reanalysis near-surface air temperature was not provided in the GSSTF products, so this was taken by the NCEP–NCAR and NCEP–DOE reanalyses for GSSTF2 and GSSTF2b, respectively.

Shie (2010) pointed out that the GSSTF2b set-1 latent heat flux trend was mainly caused by a trend in SSM/I brightness temperatures used to retrieve the bottom 500-m column water vapor. This was found to be due to temporal variations of the Earth incidence angle of the individual SSM/I satellites (Shie and Hilburn 2011). Thus, adjusted brightness temperatures are being used in an updated version (GSSTF2c) in which the temporal trends in specific humidity and latent heat flux are considerably reduced. This version is scheduled to be released in autumn of 2011. Soon, the GSSTF team will start production on a newer version (GSSTF3) that will include the adjusted brightness temperatures used in GSSTF2c plus increased horizontal (0.25° × 0.25°) and temporal (12-hourly) resolution as well as SSTs based on AMSR-E and TRMM measurements. GSSTF3 is expected to be released no later than spring of 2012.

e. J-OFURO

J-OFURO (http://dtsv.scc.u-tokai.ac.jp/j-ofuro/index.html) is produced by the School of Marine Science and Technology at Tokai University in Japan. In version 2 used here, wind speed for the sensible and latent heat fluxes is taken from SSM/I, *ERS-1* and *ERS-2*, the active microwave instrument (AMI), QuikSCAT, AMSR-E, and TMI. SST is taken from the JMA Merged Satellite and In situ Data Global Daily SST (MGDSST). Specific humidity at 10 m is retrieved from SSM/I measurements following Schlüssel et al. (1995), excluding certain data as in Schulz et al. (1993). Near-surface air temperature was taken from NCEP–DOE. Latent and sensible heat fluxes were calculated using the Coupled Ocean–Atmosphere Response Experiment (COARE), version 3.0, algorithm (Fairall et al. 2003; see appendix). Wind stresses were also available beginning in August 1999, but most of the cruise data (see section 3) occur before this time. So, we have not included the J-OFURO wind stresses here.

J-OFURO's daily 1° × 1° surface fluxes and specific humidities were provided through the dataset's Web site (http://dtsv.scc.u-tokai.ac.jp/j-ofuro/index.html). SST was calculated from the surface specific humidity. Air temperature was taken from NCEP–DOE, since it is not included in J-OFURO.

f. HOAPS

HOAPS is produced by the University of Hamburg and the Max Planck Institute for Meteorology in Germany. Version 3 (Andersson et al. 2010) compared here uses a neural network to obtain wind speed (Krasnopolsky

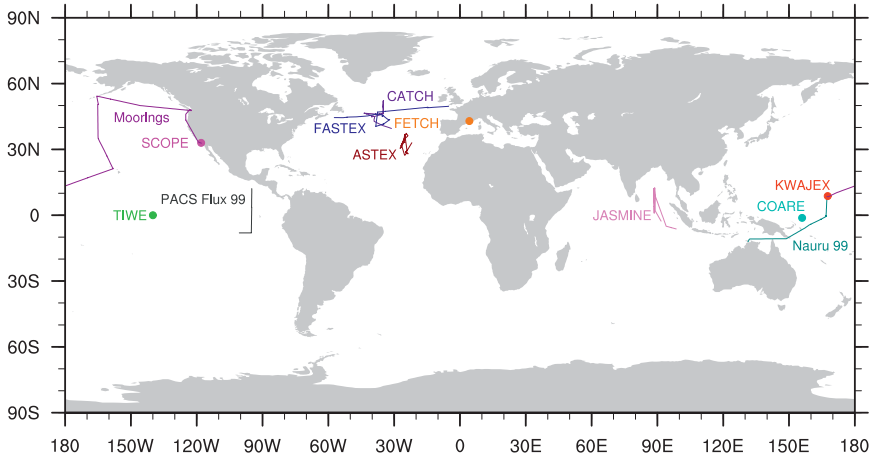

FIG. 1. The ship trajectories of each of the 12 cruises used in this study [as in Fig. 1 in Brunke et al. (2003), but repeated here for ease of reference].

et al. 1995). Specific humidity at 10 m is derived from SSM/I measurements using the updated coefficients of Bentamy et al. (2003), and surface specific humidity is derived using the Magnus formula, as described in Murray (1967), with a salinity correction. SST is taken from the Advanced Very High Resolution Radiometer (AVHRR) Oceans Pathfinder SST product (Casey et al. 2010). Near-surface air temperature is taken as the average of two methods: from the air specific humidity assuming that the relative humidity is a constant 80% (Liu et al. 1994) and from the SST assuming a constant surface–air temperature difference of 1 K (Wells and King-Hele 1990). Latent and sensible heat fluxes are computed using the COARE 3.0 algorithm (Fairall et al. 2003; see appendix).

The twice-daily (12 hourly) surface fluxes and bulk variables on a 1° × 1° grid (referred to as the HOAPS-G dataset) were downloaded from the Climate and Environmental Data Retrieval and Archive Web site (http://cera-www.dkrz.de/CERA/) at the Max Planck Institute for Meteorology.

g. *OAFlux*

OAFlux is produced by the WHOI. OAFlux is unique, in that it combines bulk variables derived from satellites with those from reanalyses (Yu and Weller 2007). Satellite-derived wind speeds come from SSM/I measurements using the Wentz (1997) algorithm and from AMSR-E and QuikSCAT. Specific humidity at 10 m is derived from SSM/I measurements using Chou et al. (1995, 1997), which is brought down to 2 m using the COARE 3.0 algorithm (Fairall et al. 2003). Satellite SST comes from NCEP's OI product (Reynolds et al. 2007). Also ingested are the corresponding bulk variables from NCEP–NCAR, NCEP–DOE, and ERA-40 (Yu et al. 2008). The latent and sensible heat fluxes are calculated using the COARE 3.0 algorithm (Fairall et al. 2003; see appendix).

The daily 1° × 1° fluxes and bulk variables were downloaded from the OAFlux project Web site (http://oaflux.whoi.edu/).

3. Ship cruises

The fluxes and bulk variables are compared with observational data during 12 cruises shown in Fig. 1. Ten of these [the Atlantic Stratocumulus Transition Experiment (ASTEX), the COARE, the Fronts and Atlantic Storm Track Experiment (FASTEX), the Joint Air–Sea Monsoon Experiment (JASMINE), the Kwajalein Experiment (KWAJEX), a cruise to service buoys in the North Pacific (Moorings), Nauru '99, the Pan-American Climate Study Flux 1999 cruise (PACS Flux '99), the San Clemente Ocean Probing Experiment (SCOPE), and the Tropical Instability Wave Experiment (TIWE)] were carried out by the NOAA Environmental Technology Laboratory (ETL) [now part of the Earth System Research Laboratory (ESRL)] between 1991 and 1999. The other two [Couplage avec l'Atmosphère en Conditions Hivernales (CATCH) in January–February 1997 and Flux, État de la Mer et Télédétection en Condition de Fetch Variable (FETCH) in March–April 1998] were operated by the Centre d'Etude des Environments Terrestre et Planétaires (CETP) [now part of the Laboratoire Atmosphères, Milieux, Observations Spatiales (LATMOS)] at L'Institut Pierre-Simon Laplace in France. For more information on these experimental cruises, please refer to Brunke et al. (2003) and the references therein.

The observed fluxes used herein are those derived from the covariance and inertial-dissipation methods.

Both are provided for sensible and latent heat fluxes as well as wind stress for all of the ETL cruises. Both were also provided for sensible and latent heat fluxes from FETCH, while only inertial-dissipation fluxes were provided for all fluxes during CATCH and for wind stress during FETCH. The covariance fluxes have been shown to be more reliable for latent and sensible heat fluxes (Fairall et al. 1996; Pedreros et al. 2003), so we use those whenever possible in this study. Inertial-dissipation wind stresses are always used, and the inertial-dissipation latent and sensible heat fluxes are used whenever covariance fluxes are unavailable. Additionally, flow distortion, ship motions, and environmental conditions have all been accounted for as described in Brunke et al. (2003).

4. Methodology

Since the temporal resolution of these flux products ranges from hourly to daily, mainly the daily average fluxes and bulk variables are compared to each other here. This also allows the bulk variables from some of the reanalyses (e.g., NCEP–NCAR and NCEP–DOE), which are only provided instantaneously, to be compared to the average quantities in the other products. It is important to point out that the daily quantities provided by the satellite-derived products may not be truly averages but a composite of all of the available passes during the day. Also, the bulk variables in GSSTF2, GSSTF2b, J-OFURO, and HOAPS are derived at 10 m, so these have been brought down to 2 m using their respective algorithms.

The product algorithms are also used offline to calculate the turbulent fluxes using the observed bulk variables as input (referred to as *algorithm fluxes*). This provides an assessment of the impact of the algorithm to the flux product error. To mimic the model timestepping used to produce the reanalyses, the observed temporal resolution was used to produce the algorithm fluxes from the model algorithms, whereas (either daily or 12 hourly) mean quantities of the observations were inputted into the algorithms used to produce the other products. Also, since wave information is not provided for these cruises, the wave age dependence in the additional roughness due to waves, as was done to produce the ECMWF reanalyses using a wave model (e.g., ECMWF 2007), was ignored here, and a constant 0.018 was used instead (see appendix). There would be a small impact on the algorithm fluxes from the ECMWF algorithm, especially for conditions in which young waves are prevalent.

The crux of this study is to analyze the uncertainties associated with the total error of a product's flux. Thus, the total flux error can be divided into two uncertainties as follows:

$$F_{\mathrm{prod}} - F_{\mathrm{obs}} = (F_{\mathrm{prod}} - F_{\mathrm{algor}}) + (F_{\mathrm{algor}} - F_{\mathrm{obs}}), \quad (4)$$

where F_{prod} is the *product flux* (the flux directly from each product), F_{obs} is the *observed flux*, and F_{algor} is the algorithm flux (computed using the respective offline algorithm inputted with the observed bulk variables). The first term on the right-hand side, called the "bulk-variable-caused uncertainty," is predominately due to the difference in bulk variables, as the same algorithm was used to produce both F_{prod} and F_{algor}. There is also some uncertainty included in this term because of the possible mismatch between a point measurement and a gridbox average. However, this should be lessened by the temporal averaging that is applied to the data from a moving ship. The last term is called the "residual uncertainty" and is partly due to the uncertainty caused by problems in the algorithm, as discussed in Zeng et al. (1998) and Brunke et al. (2002, 2003). There is also some measurement uncertainty in this term, as mean covariance latent heat flux, covariance sensible heat flux, and inertial-dissipation wind stress typically have an uncertainty of 4 W m^{-2}, 2 W m^{-2}, and 5%, respectively (Fairall et al. 1996).

In Brunke et al. (2003), we found it helpful to rank bulk flux algorithms to assess which ones were least problematic. Thus, here we employ a similar strategy to assess the overall quality of the ocean surface fluxes in the products presented here by ranking them for each flux based upon their total bias (i.e., average error, $\overline{F_{\mathrm{prod}} - F_{\mathrm{obs}}}$) and their SDEs.

Thus, for each cruise, a score ($s^i_{\mathrm{bias}_F}$) from 1 for the lowest bias magnitude to 11 (8 for wind stress) for the highest bias magnitude is assigned to each product's flux for each cruise. A similar score is also assigned to each product's flux for each cruise based upon their SDEs ($s^i_{\mathrm{SDE}_F}$). This method assigns equal weighting to each cruise with variable amounts of valid points (see Table 3 in Brunke et al. 2003). So, scores are also assigned to each product's flux based upon the all-cruise biases (S_{bias_F}) and SDEs (S_{SDE_F}), as presented in Table 1. An overall ranking of each product's flux is, then, the average of these four values as follows:

$$S_F = \frac{1}{4}(\overline{s_{\mathrm{bias}_F}} + \overline{s_{\mathrm{SDE}_F}} + S_{\mathrm{bias}_F} + S_{\mathrm{SDE}_F}). \quad (5)$$

5. Results

a. Evaluation of the total biases

Most of the bulk algorithms are not well tested at high wind speeds, since few observations are made in this regime. Zeng et al. (1998) found that the algorithm

TABLE 1. The bias ($\overline{F_{prod} - F_{obs}}$), mean bulk variable uncertainty ($\overline{F_{prod} - F_{algor}}$), mean residual uncertainty ($\overline{F_{algor} - F_{obs}}$), and SDE for each product during all cruises based on the daily means. Those based on just CATCH and FASTEX are given in parentheses. The boldface numbers are the higher values between $\overline{F_{prod} - F_{algor}}$ and $\overline{F_{algor} - F_{obs}}$.

Product	$\overline{F_{prod} - F_{obs}}$	$\overline{F_{prod} - F_{algor}}$	$\overline{F_{algor} - F_{obs}}$	SDE
		LH flux (W m^{-2})		
MERRA	2.6 (−15.6)	**3.3 (−14.6)**	−0.7 (−1.0)	31.2 (42.4)
ERA-40	16.7 (8.7)	3.3 (0.2)	**13.4 (8.4)**	37.5 (50.4)
ERA-Interim	17.6 (8.3)	3.8 (0.8)	**13.8 (7.5)**	34.7 (45.8)
NCEP–NCAR	11.2 (11.3)	−13.4 (−29.8)	**24.5 (41.1)**	41.7 (65.6)
NCEP–DOE	20.2 (38.1)	−4.3 (−3.1)	**24.5 (41.1)**	46.1 (72.9)
CFSR	19.3 (2.9)	8.2 (−12.8)	**11.0 (15.7)**	44.8 (77.5)
GSSTF2	9.0 (31.7)	1.6 **(36.6)**	**7.0** (−4.8)	49.6 (76.9)
GSSTF2b	−0.6 (25.6)	−8.0 **(30.5)**	**7.0** (−4.8)	44.9 (71.1)
J-OFURO2	−3.0 (8.3)	**−10.6 (6.4)**	7.3 (1.9)	45.4 (70.6)
HOAPS	1.7 (−9.8)	−5.9 **(−8.6)**	**7.3** (1.9)	50.3 (85.3)
OAFlux	11.6 (24.7)	4.2 **(22.8)**	**7.3** (1.9)	41.0 (64.9)
		SH flux (W m^{-2})		
MERRA	−0.8 (−14.5)	**−1.0** (−0.3)	0.2 **(−14.2)**	14.8 (30.2)
ERA-40	0.9 (−9.0)	**0.9** (6.5)	0.1 **(−15.5)**	18.7 (41.5)
ERA-Interim	2.7 (−4.5)	**2.6** (11.3)	0.1 **(−15.8)**	14.2 (32.4)
NCEP–NCAR	6.0 (3.1)	−9.0 (−16.6)	**15.0 (19.7)**	23.2 (54.5)
NCEP–DOE	0.7 (1.9)	−14.3 (−17.8)	**15.0 (19.7)**	22.1 (53.6)
CFSR	−0.3 (−15.3)	**−12.0 (−23.3)**	11.7 (8.1)	22.6 (53.0)
GSSTF2	4.4 (−6.0)	**5.2** (12.1)	0.1 **(−17.3)**	20.1 (43.3)
GSSTF2b	−0.4 (−9.2)	**0.4** (8.9)	0.1 **(−17.3)**	18.5 (40.6)
J-OFURO2	2.3 (0.0)	**−1.9 (13.0)**	1.0 (−13.0)	20.1 (43.6)
HOAPS	−1.4 (−20.3)	**−1.6** (−3.8)	1.0 **(−13.0)**	19.0 (35.4)
OAFlux	2.2 (4.7)	**1.3 (17.7)**	1.0 (−13.0)	18.1 (43.1)
		Wind stress (10^{-3} N m^{-2})		
MERRA	−6.2 (−25.8)	1.0 (14.4)	**−7.2 (−40.2)**	42.3 (71.7)
ERA-40	−3.9 (−8.3)	**−3.4** (3.2)	−0.6 **(−11.5)**	55.1 (89.9)
ERA-Interim	−2.3 (−2.5)	**−2.2** (12.4)	−0.1 **(−14.9)**	43.3 (69.9)
NCEP–NCAR	−7.6 (−1.0)	−3.7 (19.9)	**−3.9 (−20.9)**	48.5 (86.1)
NCEP–DOE	−0.1 (21.1)	3.8 **(42.0)**	**−3.9** (−20.9)	50.4 (98.6)
CFSR	4.8 (33.4)	**8.2 (55.0)**	−3.5 (−21.5)	89.1 (214.6)
GSSTF2	4.7 (21.6)	**14.5 (58.3)**	−8.0 (−34.9)	106.9 (229.4)
GSSTF2b	7.9 (38.8)	**17.7 (75.6)**	−8.0 (−34.9)	92.7 (204.0)

fluxes—that is, the fluxes calculated by bulk algorithms using observed bulk variables as input—diverge for high wind speeds. Figure 2a shows that the suite of cruise data in this study covers a wide range of wind speeds, from very low (<1 m s^{-1}) to ≥10 m s^{-1}. Ten of the 12 cruises, most of them in the tropics and subtropics, have the most number of data points in the moderate range (from 3 to 6 m s^{-1}). The other two (CATCH and FASTEX) have the most number of points in the high wind speed range, and FETCH has many high wind speed points as well as many in the moderate range.

Thus, to examine how well the data products produce fluxes at high wind speeds, the quality of the daily mean fluxes from CATCH and FASTEX are compared with that of all other cruises by comparing the product fluxes with the observed fluxes in Figs. 3–5. For LH and SH fluxes (Figs. 3, 4), the reanalysis regression slopes (except for CFSR LH flux) are generally higher for CATCH/FASTEX than for the other cruises, whereas some of the slopes of the regressions for the satellite-derived products are lower during CATCH/FASTEX. Except for NCEP–DOE's LH flux, NCEP–NCAR's and NCEP–DOE's regression slopes are usually very close to 1, while the scatter in these two products as well as CFSR is higher than the other products. The wind stresses from the reanalyses are better than those from the GSSTF products, especially GSSTF2 (Fig. 5). As for LH and SH fluxes, the wind stress regression slopes for the reanalyses, except CFSR, are all higher for CATCH/FASTEX. CFSR and both versions of GSSTF2 overestimate low wind stresses and underestimate high wind stresses, particularly during the cruises other than CATCH and FASTEX (Fig. 5).

The product total biases based on the daily means for all cruises are presented in Table 1. For LH flux, total biases range from −3.0 W m^{-2} for J-OFURO to 20.2 W m^{-2} for NCEP–DOE. The highest total biases come from the

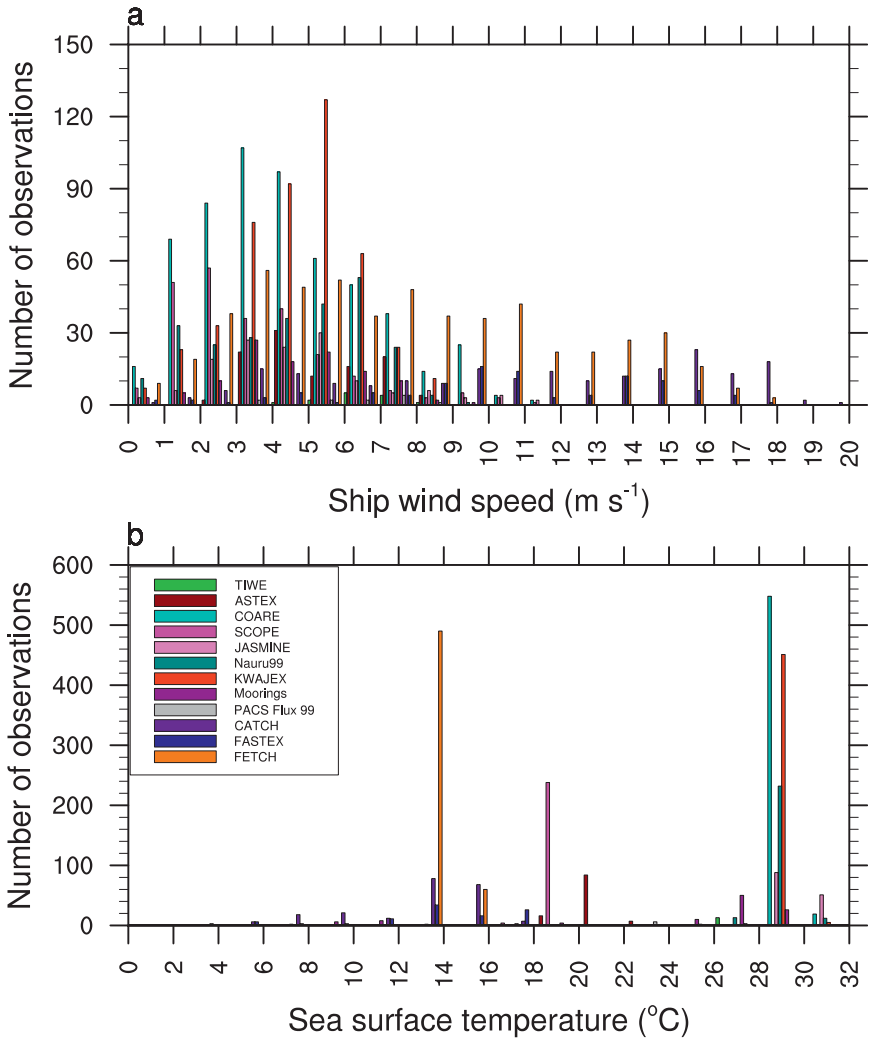

FIG. 2. The number of (a) wind speed observations for each of the 12 cruises for every 1 m s^{-1} bin and (b) SST observations for every 2°C bin for each of the 12 cruises.

reanalyses other than MERRA. SH fluxes over ocean are generally much smaller in magnitude, so the total biases are generally smaller as well (from -1.4 W m^{-2} for HOAPS to 6.0 W m^{-2} for NCEP–NCAR). For wind stress, the lowest total bias comes from NCEP–NCAR (-7.6×10^{-3} N m^{-2}), while GSSTF2b has the highest bias (7.9×10^{-3} N m^{-2}).

Also shown in Table 1 are the SDEs based on the daily means. HOAPS has the highest SDE for LH and SH fluxes (50.3 and 26.8 W m^{-2}, respectively). The lowest LH flux SDEs come from MERRA, ERA-40, and ERA-Interim, while GSSTF2, J-OFURO, and HOAPS have higher SDEs than the other products.

The all-cruise biases and SDEs are dominated by lower wind speed data points, so the biases and SDEs for just CATCH and FASTEX are also shown in parentheses in Table 1. The SDEs from these two cruises alone are higher than for all the cruises combined, consistent with the higher scatter seen in Figs. 3–5. Most products have biases that are higher in magnitude for CATCH/FASTEX.

The CATCH/FASTEX biases in Table 1 provide a hint as to the effect of wind speed on the product fluxes. To get a better understanding of this, the total biases for each product averaged for 1 m s^{-1} bins are shown in Figs. 6a–c. The product biases are generally close to 0 in the moderate wind speed range but diverge at very low wind speeds (<1 m s^{-1}) and high wind speeds (≥ 9 m s^{-1}). These regimes also have a low number of total observations (Fig. 2a), and there are higher standard deviations in the observed fluxes in the high wind speed bins (not shown).

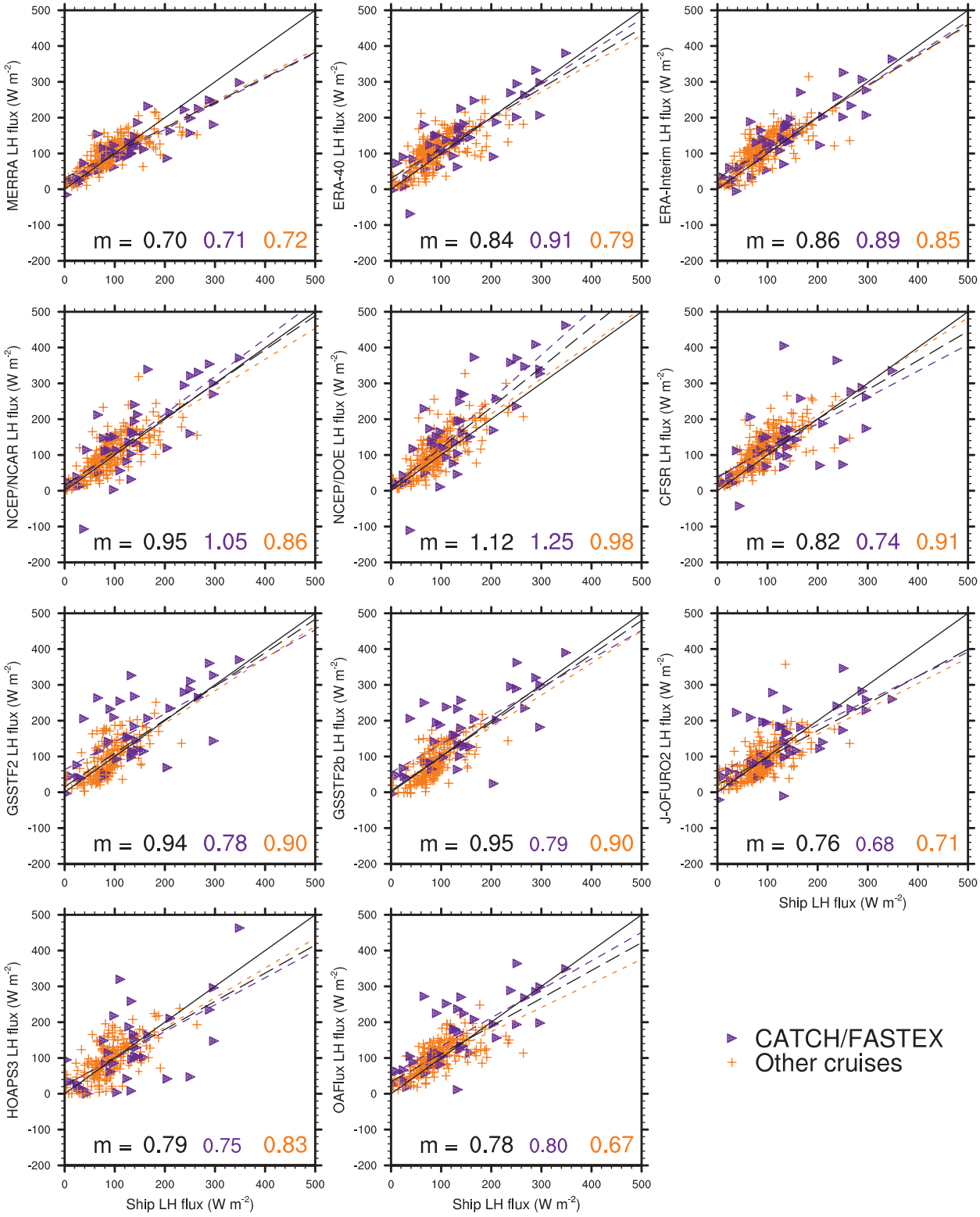

FIG. 3. Daily mean product LH flux compared with observed LH flux from CATCH and FASTEX (purple triangles) and from every other cruise (orange crosses). The regression lines and their slopes (m) in their corresponding colors for all of the data (black dashed line), CATCH and FASTEX only (purple short-dashed line), and the other cruises (orange dotted line) are also given. The solid black line is the 1-to-1 line.

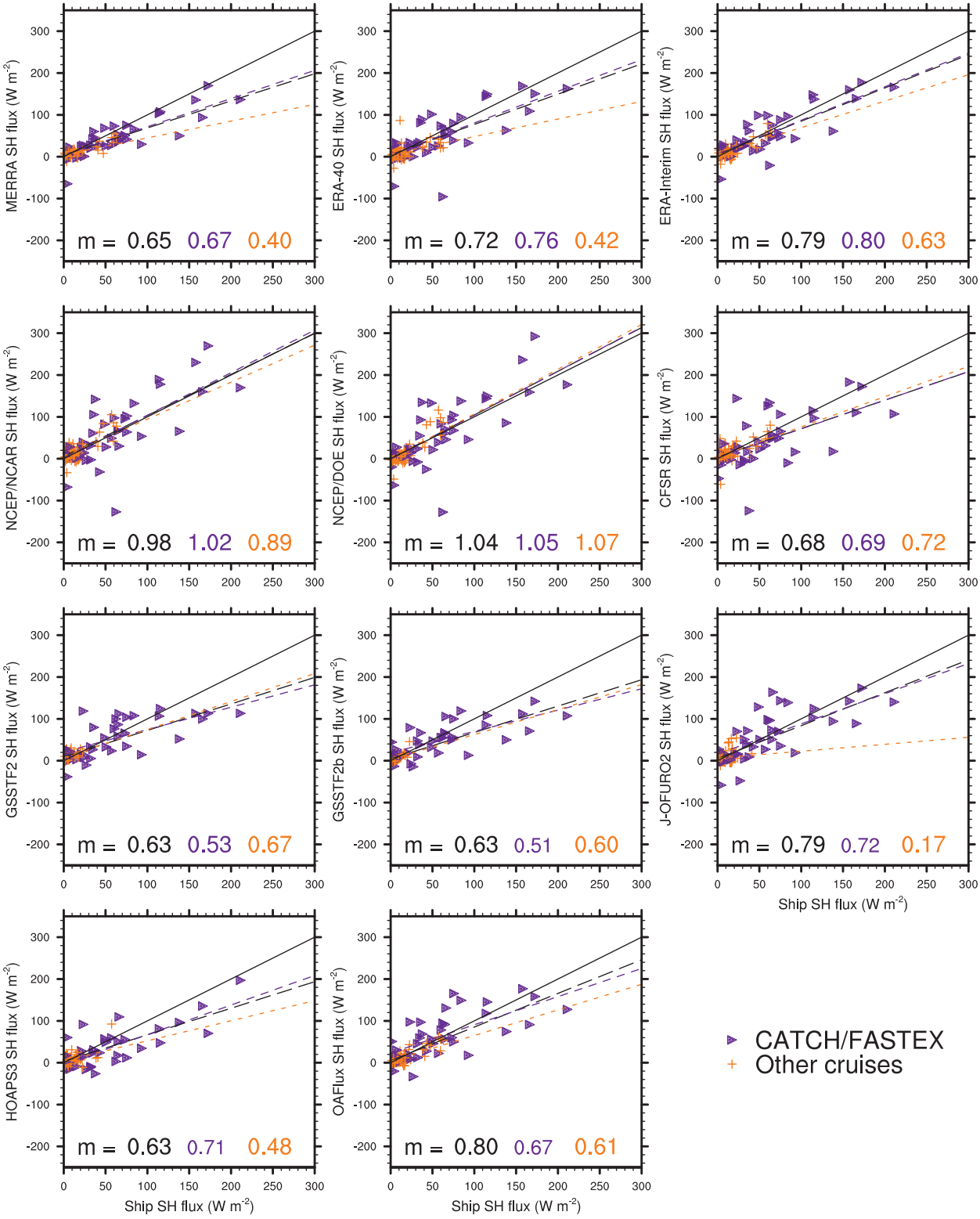

FIG. 4. As in Fig. 3, but for SH flux.

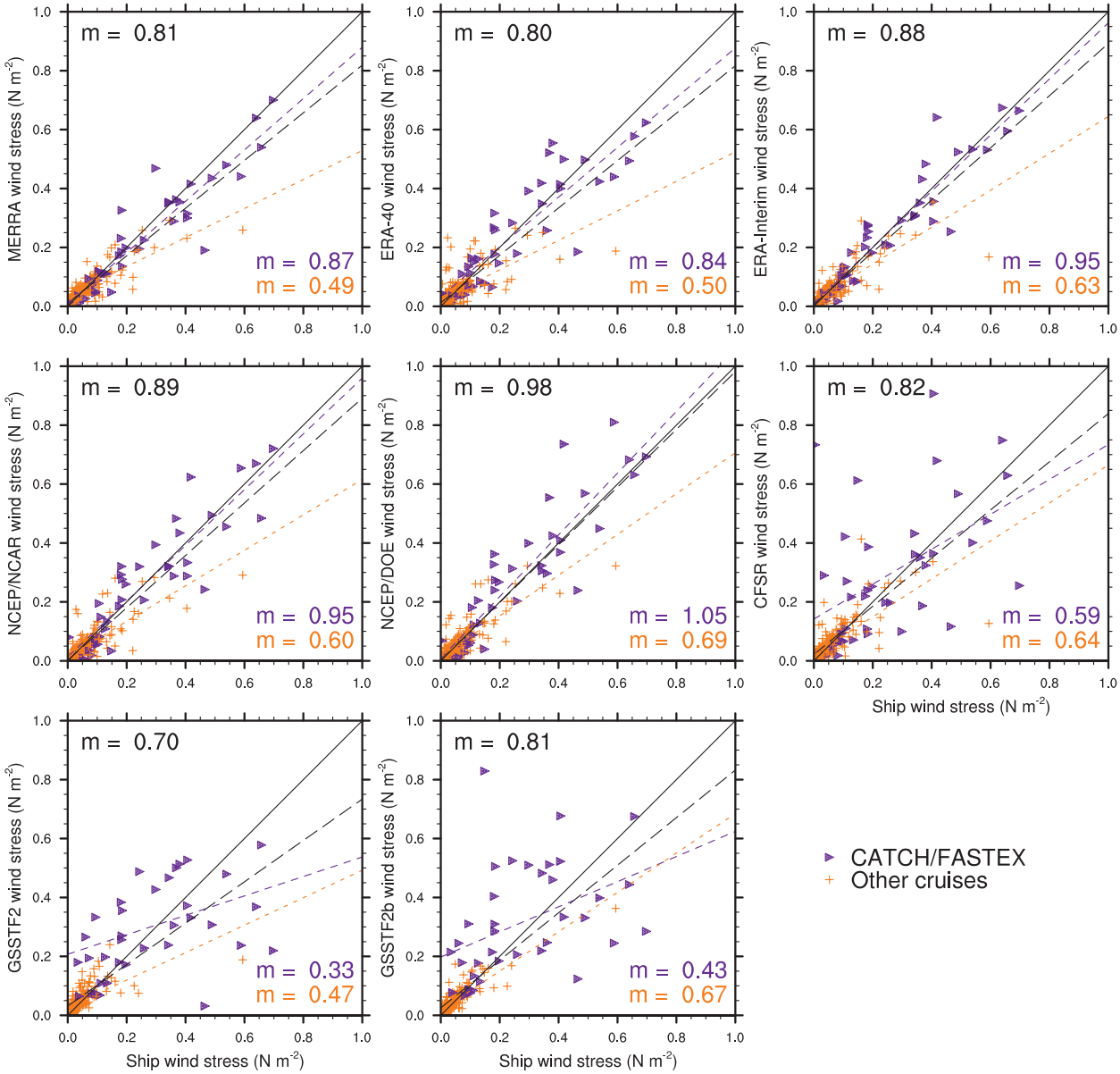

FIG. 5. As in Fig. 3, but for wind stress.

Another way to differentiate between data from the various cruises is by SST. Tropical SSTs are warm with most temperatures > 26°C, while the midlatitude cruises have cold SSTs with most SSTs lying between 12° and 22°C (Fig. 2b). Figures 7a–c present the total product biases averaged over 2°C bins. The total biases in SH flux and wind stress from the products are typically around 0 except for SSTs < 18°C (Figs. 7b,c). These low SSTs correspond to the midlatitude cruises of CATCH and FASTEX as well as some other midlatitude and subtropical cruises (Fig. 2b), whereas the total biases in product LH flux do not show a strong dependence on SST (Fig. 7a). As was seen for high wind speeds, the scatter in the observed fluxes is higher for SSTs of <18°C (not shown).

b. *Assessment of the general performance of the data products*

Based upon the scores obtained from Eq. (6), we put the products into categories of performance for each of the fluxes. For the fluxes that all products have (LH and SH flux), there are three categories: A for the four "best performing" or lowest overall scores (S_F), C for the

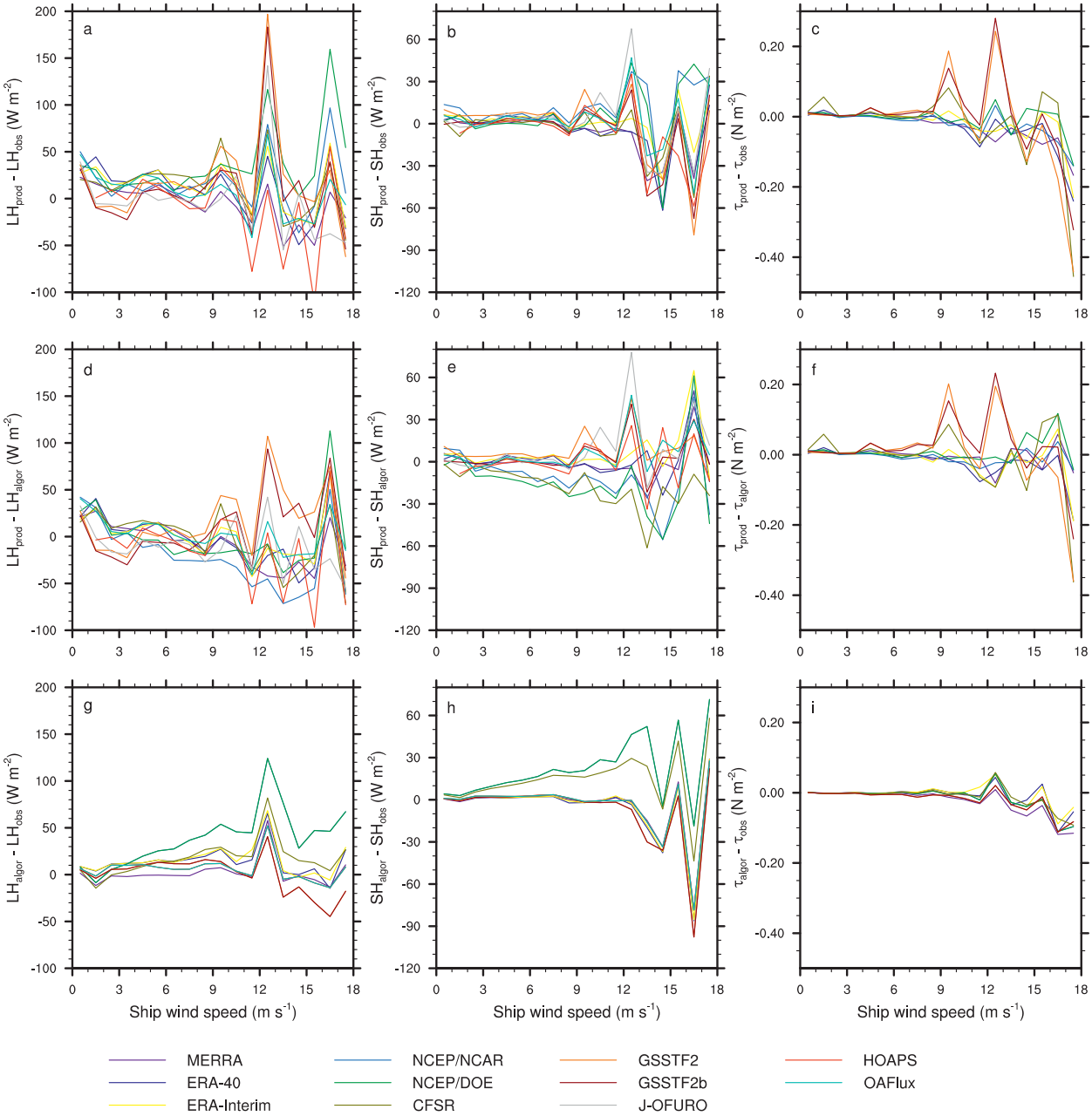

FIG. 6. (a)–(c) Bias ($F_{prod} - F_{obs}$), (d)–(f) mean bulk variable uncertainty ($F_{prod} - F_{algor}$), and (g)–(i) mean algorithm uncertainty ($F_{algor} - F_{obs}$) based on the daily means for every 1 m s^{-1} wind speed bin in (left) LH flux, (center) SH flux, and (right) wind stress for each product over all 12 cruises.

three "worst performing" or highest S_F, and B for the other four. Because wind stress is only provided by eight of the products, there are only two categories: A and B. This is presented in Table 2, in which the products are listed in alphabetical order per category.

Of the reanalyses, MERRA is in category A for all three fluxes. ERA-Interim falls into category A for two fluxes (LH flux and wind stress), while ERA-40 is also in category A for LH flux. NCEP–DOE is in category A for two fluxes (SH flux and wind stress).

The satellite-derived products generally do fairly well, falling into either category A or B for LH and SH fluxes. For instance, GSSTF2b is in category A for both LH and SH fluxes. Conversely, GSSTF2 LH and SH fluxes are in category C, and HOAPS and J-OFURO have category C LH and SH fluxes, respectively.

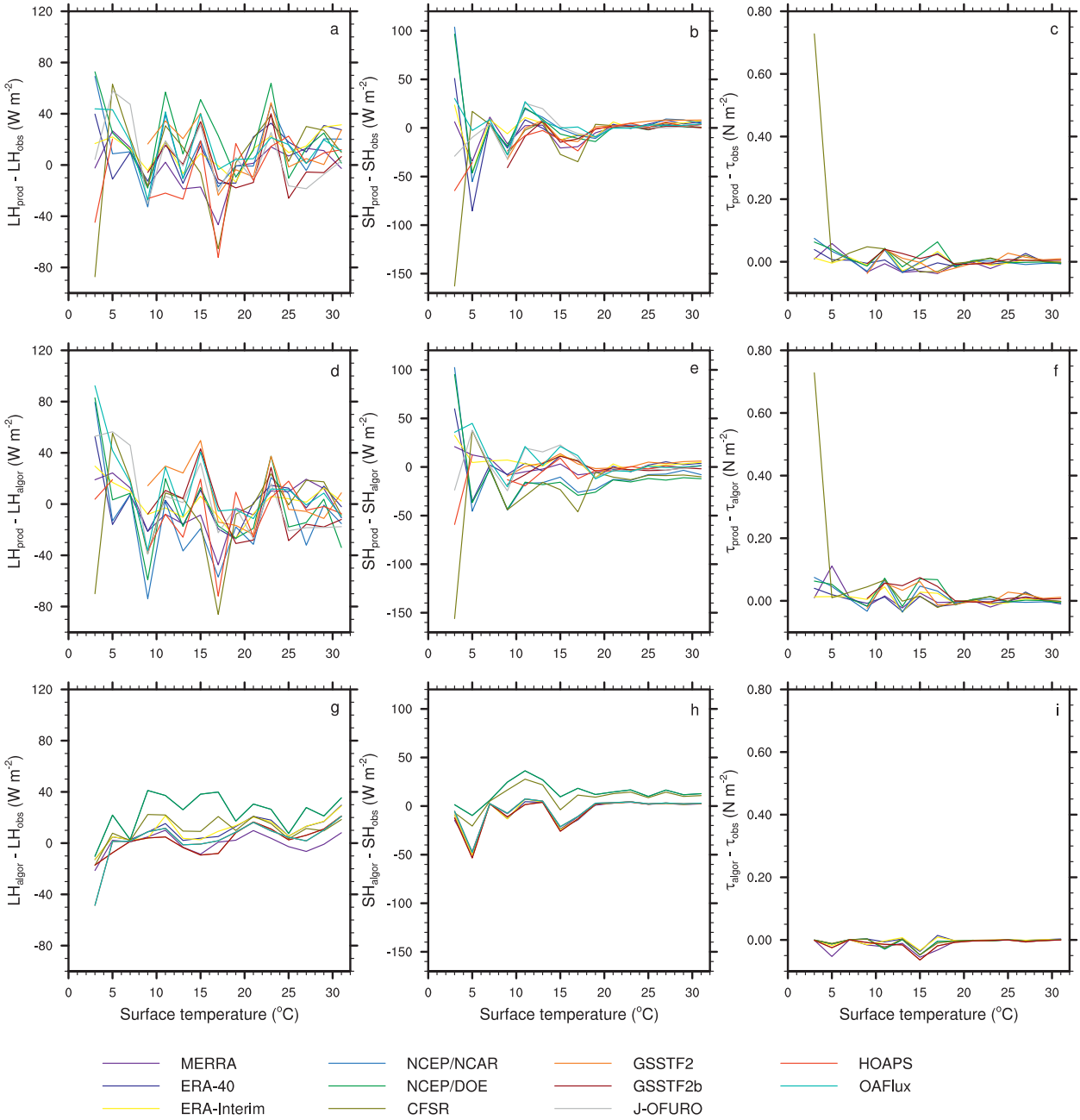

FIG. 7. As in Fig. 6, but for 2°C bins of SST.

c. *Evaluation of the bulk variables*

To understand the total biases in Figs. 3–7 and Table 1 and the general levels of performance in Table 2, the contributions to the biases are examined by first looking at the accuracy of the bulk variables (wind speed, near-surface air temperature, SST, and near-surface air humidity) that are used as input into the bulk algorithms.

First, Fig. 8a compares the biases in 2-m specific humidity from all of the products for each cruise. The all-cruise biases in this quantity are quite low (from −0.5 to 0.9 g kg^{-1}), which is because this quantity is overestimated in some cruises and underestimated in others (Fig. 8a). During some cruises, the difference between the product and observed specific humidity can be quite substantial. For instance, six products (NCEP–NCAR, NCEP–DOE, GSSTF2, GSSTF2b, J-OFURO, and

TABLE 2. The general level of performance of the 11 data products compared here based on Eq. (5) using all of the cruise data. The products are put into one of three categories: A, B, or C. Products in each category are listed in alphabetical order.

Group	LH flux	SH flux	Wind stress
A	ERA-40	GSSTF2b	ERA-40
	ERA-Interim	MERRA	ERA-Interim
	GSSTF2b	NCEP–DOE	MERRA
	MERRA	OAFlux	NCEP–DOE
B	CFSR	CFSR	CFSR
	J-OFURO	ERA-40	GSSTF2
	NCEP–NCAR	ERA-Interim	GSSTF2b
	OAFlux	HOAPS	NCEP–NCAR
C	HOAPS	GSSTF2	
	GSSTF2	J-OFURO	
	NCEP–DOE	NCEP–NCAR	

OAFlux) have biases $> +1$ g kg^{-1} during TIWE (Fig. 8a). OAFlux's overestimation at this location is due to its contribution from NCEP–NCAR and NCEP–DOE and only one of the ECMWF reanalyses (ERA-40), which has a better representation of specific humidity here. Problematic for the products that derived the near-surface specific humidity only from satellite data (J-OFURO and HOAPS) is SCOPE; all of these products underestimate this quantity in excess of ~1.8 g kg^{-1}. The location of this cruise is within ~70 km of the Southern California coast, so the proximity near the coast may have an effect here.

All-cruise biases are higher (-2.1 to 1.0 m s^{-1}) for 10-m wind speed. Seven products (NCEP–DOE, CFSR, GSSTF2, GSSTF2b, J-OFURO, HOAPS, and OAFlux) generally overestimate wind speed for most cruises, with the highest overestimations by the satellite-derived products (Fig. 8b). For FETCH, the reanalyses other than NCEP–DOE plus OAFlux underestimate wind speed in excess of 1 m s^{-1}.

The biases in 2-m air temperature can be quite high as well (Fig. 8c). The reanalyses all overestimate this quantity in excess of 1°C for SCOPE. These products have an inexplicable warming during the latter half of the cruise that is not seen in the observations until the very last day (not shown). J-OFURO and OAFlux also use the air temperatures from reanalyses, so they also overestimate air temperature in excess of 1°C during SCOPE (Fig. 8c). Another problematic cruise for 2-m air temperature in some products is FETCH. Particularly, NCEP–NCAR and NCEP–DOE underestimate this quantity by on average ~2.4° and ~1.8°C, respectively. Since GSSTF2, GSSTF2b, and J-OFURO also utilize these temperatures, they also underestimate this quantity substantially. The underestimation is not as great in OAFlux because it also ingests ERA-40 temperatures, which have a small overestimation during this cruise.

SST biases are generally within 1°C except for a few instances (Fig. 8d). During SCOPE, ERA-40, NCEP–NCAR, and NCEP–DOE have lower SSTs in excess of 1°C. During FETCH, NCEP–NCAR and NCEP–DOE also underestimate SSTs by more than 1°C. Since GSSTF2 and GSSTF2b use the SSTs from NCEP–NCAR and NCEP–DOE, respectively, their SSTs are also much lower than observed during this cruise.

Figure 9 shows the SDEs of the product bulk variables. The satellite-derived 2-m specific humidities in GSSTF2, GSSTF2b, J-OFURO, and HOAPS have slightly higher SDEs than those from the reanalyses and OAFlux (Fig. 9a). Wind speed SDEs, especially from the products that use satellite-derived quantities, are usually higher during CATCH, FASTEX, FETCH, and Moorings, which all experienced some higher wind speeds at higher latitudes (Fig. 9b). SDEs for 2-m air temperature are generally small (<1°C) except for a few instances (Fig. 9c). SST SDEs are generally even smaller (<0.5°C) except during PACS Flux 99; CATCH; FASTEX; and Moorings for all but GSSTF2, J-OFURO, and OAFlux (Fig. 9d).

d. Evaluation of the uncertainties

The combined effect of the bulk variable biases seen in the last subsection is expressed in the bulk-variable-caused uncertainties given in the third column in Table 1. For LH and SH fluxes, the all-cruise mean bulk-variable-caused uncertainties range from very high negative values to more modest positive values (from -13.4 W m^{-2} from NCEP–NCAR to 8.2 W m^{-2} from CFSR for LH flux and from -14.3 W m^{-2} from NCEP–DOE to 5.2 W m^{-2} from GSSTF2 for SH flux). The lowest mean wind stress bulk-variable-caused uncertainty is from NCEP–NCAR (-3.4×10^{-3} N m^{-2}), while the highest values are from GSSTF2 and GSSTF2b (1.5×10^{-2} and 1.7×10^{-2} N m^{-2}, respectively).

How much these contribute to the total biases can be assessed by comparing them with the residual uncertainties in the fourth column of Table 1. The larger contributor to the total bias depends on the product. In some, the residual uncertainty contributes more to the total bias, while others have the bulk-variable-caused uncertainty contributing more. For LH flux, all products—except MERRA, J-OFURO, and HOAPS—have the most uncertainty coming from the residual uncertainty. In contrast, the total SH flux biases of nine products (MERRA, ERA-40, ERA-Interim, CFSR, GSSTF2, GSSTF2b, J-OFURO, HOAPS, and OAFlux) are predominately composed of the bulk-variable-caused uncertainties. For wind stress, five products (ERA-40, ERA-Interim, CFSR, GSSTF2, and GSSTF2b) have the bulk-variable-caused uncertainties contributing the most, while the residual

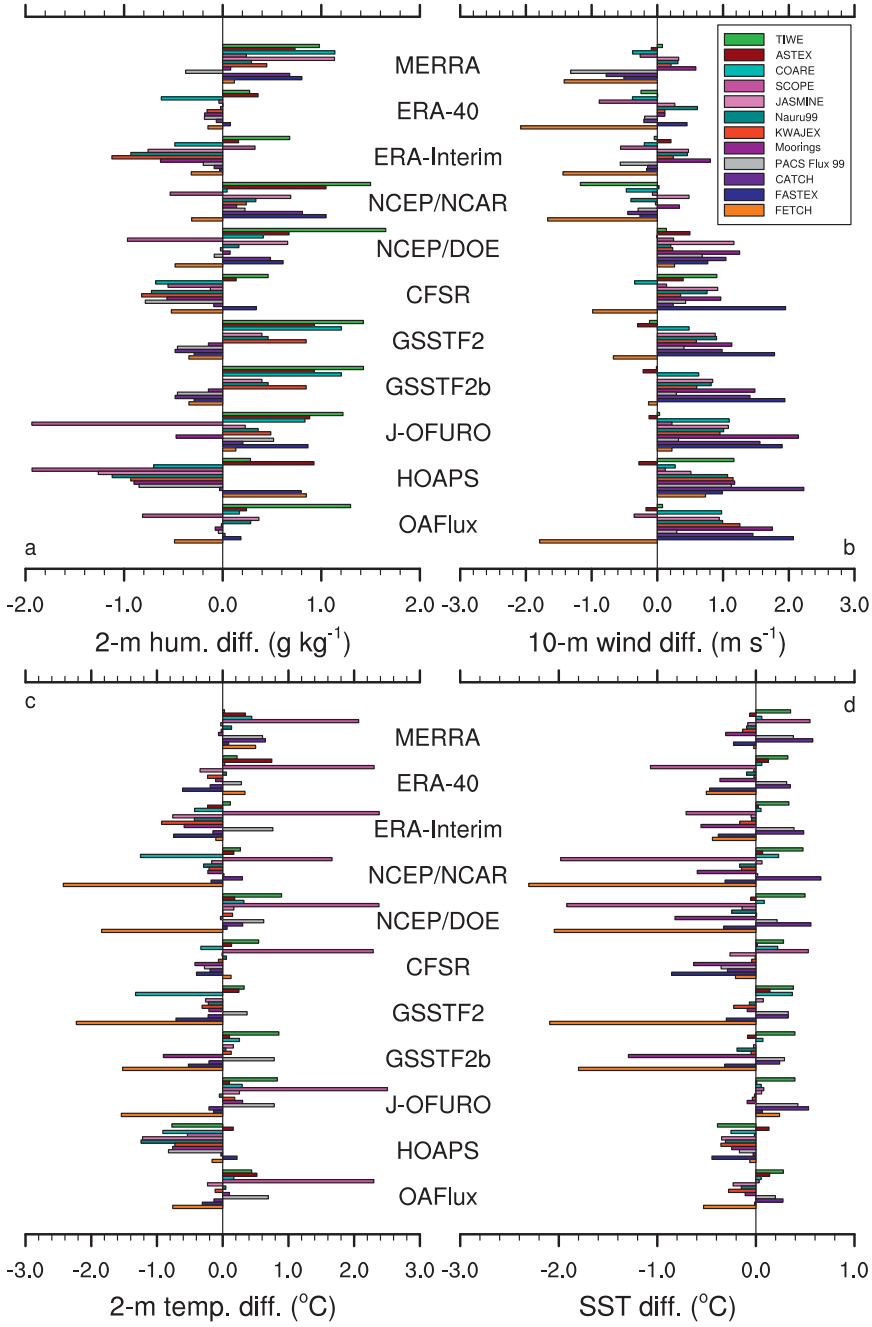

FIG. 8. The mean difference between the daily mean product and observed (a) 2-m specific humidity, (b) 10-m wind speed, (c) 2-m temperature, and (d) SST for each cruise. Note that there are no valid points from GSSTF2 and GSSTF2b for SCOPE.

uncertainties predominate in the total biases from MERRA, NCEP–NCAR, and NCEP–DOE.

Even though the residual uncertainties are composed of contributions from the algorithm and measurement uncertainties, they can still be used to get some understanding of the algorithm uncertainties. For instance, we can see a lessening of the residual uncertainty in CFSR compared to NCEP–NCAR and NCEP–DOE because of the use of the Zeng et al. (1998) roughness lengths for heat and moisture in CFSR. Also, the LH flux residual uncertainties from the satellite-derived and combined products are all lower than those from the ECMWF and NCEP reanalyses, and the SH flux residual uncertainties from these products are lower than those from the NCEP

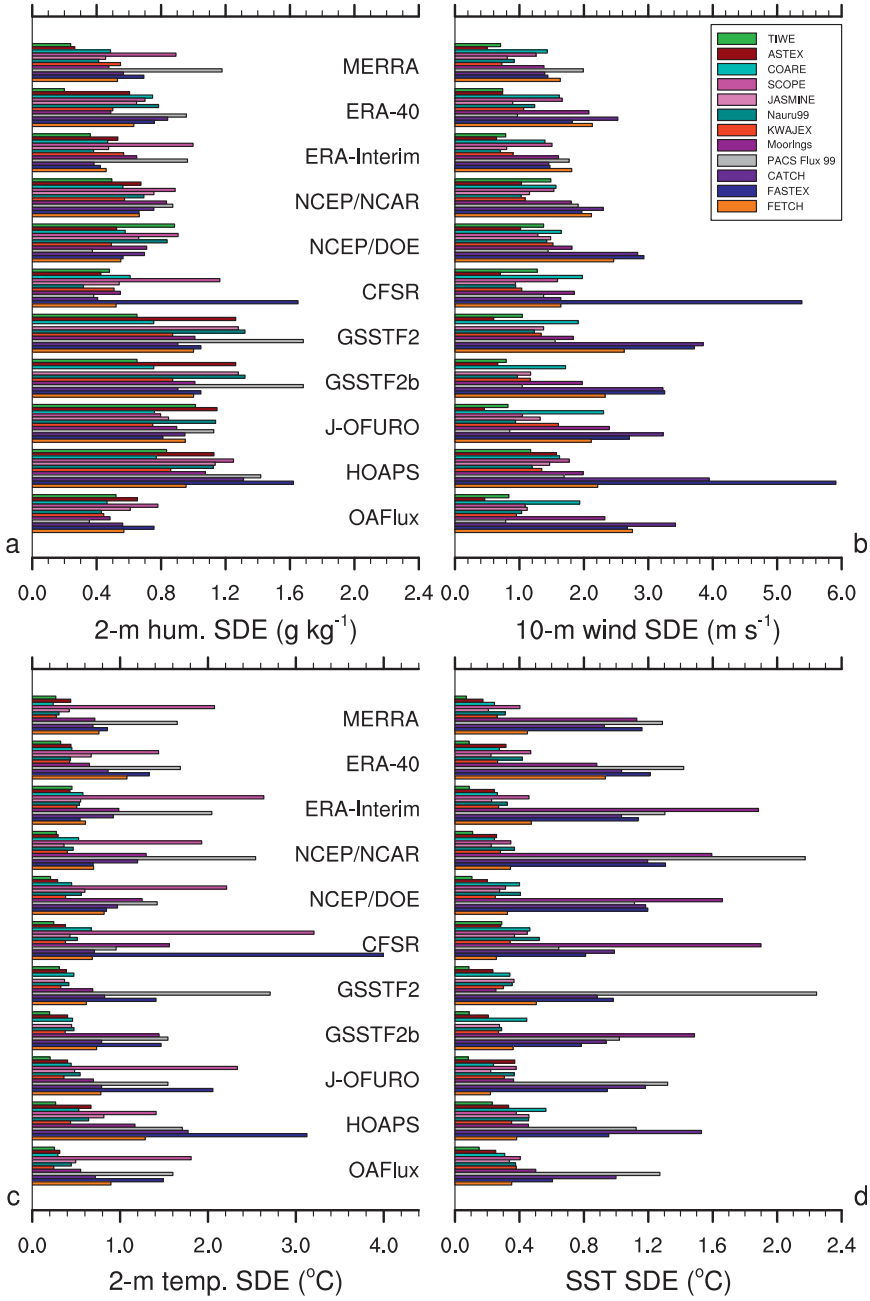

FIG. 9. As in Fig. 8, but for SDEs.

reanalyses. J-OFURO, HOAPS, and OAFlux use the COARE 3.0 algorithm, which Brunke et al. (2003) found to be one of the least problematic overall.

The values for just CATCH and FASTEX are shown in parentheses in the third and fourth columns of Table 1. Different from the tropical- and subtropical-dominated all-cruise means, the reanalysis means during CATCH/FASTEX are usually dominated by the residual means, whereas the satellite-derived means (including OAFlux) are usually dominated by the bulk-variable-caused uncertainties.

An idea of whether these uncertainties are systematic or cruise dependent can be gleaned from the middle and bottom rows of Figs. 6, 7. Similar to the total biases, the spread in the mean bulk-variable-caused and residual uncertainties is greatest at high wind speeds (>10 m s^{-1}), but the spread is higher in the bulkvariable-caused uncertainties in this regime (Figs. 6d–i). Also, the NCEP

TABLE 3. The total biases and SDEs of various temporal averages of product LH flux for data from all of the cruises.

Product	Average				
	1 h	3 h	6 h	12 h	Day
		Total bias			
MERRA	0.1	1.9	2.5	0.2	2.6
CFSR	17.5	19.3	20.9	19.3	19.3
ERA-Interim		17.3	18.4	16.2	17.6
ERA-40			16.8	14.8	16.7
NCEP–NCAR			11.1	8.4	11.2
NCEP–DOE			19.5	16.6	20.2
		SDE			
MERRA	41.5	38.7	36.4	36.0	31.2
CFSR	53.2	51.7	49.3	48.0	44.8
ERA-Interim		41.9	38.7	37.2	34.7
ERA-40			42.8	41.3	37.5
NCEP–NCAR			46.5	44.5	41.7
NCEP–DOE			52.6	49.5	46.1

products have the highest residual uncertainties overall for wind speeds > 2 m s^{-1} (Figs. 6g–i), but the use of the Zeng et al. (1998) roughness lengths in CFSR does drastically improve these uncertainties for LH flux, making them much closer to the others at ~0 (Fig. 6g).

The middle and bottom rows of Fig. 7 present the mean uncertainties for the 2°C SST bins. The spread in the total biases of Figs. 7a–c is predominantly due to that of the bulk variable uncertainties in Figs. 7d–f. Also, the NCEP products' residual uncertainties are systematically larger than those of the other products for all SSTs, so this is not a regional or regime-specific bias.

e. The effects of temporal resolution

So far, we have not looked at the effects of temporal resolution on the data products. The most recent reanalyses (MERRA, CFSR, and ERA-Interim) have increased temporal resolution (less than 6 hourly). Table 3 presents the total biases and SDEs of various temporal means of the reanalysis LH fluxes using data from all of the cruises. The biases vary for each temporal resolution within at most ~4 W m^{-2}, so there is no systematic lessening or amplifying of the product biases with temporal resolution in LH flux nor in SH flux and wind stress (not shown). In contrast, the SDE generally decreases as the averaging period increases. This is also the case for SH flux and wind stress (not shown). Thus, any "diurnal cycle" in the reanalysis errors is smoothed out in the daily means that have been compared so far.

This smoothing may have an effect on the rankings in Table 1. The choice of the number of lowest scores in category A and the highest scores in category B is somewhat arbitrary. Any small change in the score might bump a product from one category to another. Thus, the relative position of the reanalysis rankings are tested by rescoring the six reanalyses based upon their 6-hourly means instead of their daily means. The relative positions of the reanalysis rankings are only affected for wind stress where ERA-Interim is improved over ERA-40.

Figure 10 compares the 3-hourly mean bulk variables and fluxes from the reanalyses with the highest temporal resolution (MERRA, CFSR, and ERA-Interim) with observations over the course of KWAJEX during which the ship was fairly stationary at ~8°22′N, 167°44′E. This was chosen because the diurnal cycle would be resolved by the ship measurements without it being due to the movement of the ship. Because of the stationarity, more of the differences between the reanalysis values and the observed values would be due to the comparison between point measurements and grid average values than in most of the other cruises in which the ship was underway. However, some bias could be due to either a mean overall bias or a temporal one. For instance, observed specific humidity is fairly uniform across the day at ~19 g kg^{-1}. The diurnal cycles in the reanalyses are also small, with MERRA having humidities slightly above observed within one standard deviation of observed mean humidity; CFSR and ERA-Interim have humidities that are ~1 and ~1.25 g kg^{-1}, respectively, lower, generally falling below one standard deviation of the observed mean (Fig. 10a). Also, the diurnal cycle of 2-m air temperature in MERRA and CFSR is fairly consistent with observations, while those in ERA-Interim are ~1°C lower outside one standard deviation from the observed mean (Fig. 10b).

Conversely, the SST diurnal cycle in MERRA and ERA-Interim is virtually nonexistent compared to the slight one of ~0.6°C observed (Fig. 10c). CFSR with its coupling to an ocean model has a slight diurnal cycle in SST; however, since the uppermost layer is 10 m thick, it is unable to resolve the surface skin temperature diurnal cycle (Brunke et al. 2008). Near-surface wind speeds in MERRA are fairly consistent with the observed means (Fig. 10d), with the root-mean-square error (RMSE) of the eight 3-hourly averages being 0.2 m s^{-1}. In contrast, the RMSEs of CFSR and ERA-Interim are 0.6 and 0.3 m s^{-1}, respectively.

The combined effect of these bulk variable errors on the diurnal cycle of fluxes can be seen in the bottom two rows of Fig. 10. The underestimated specific humidities in CFSR and ERA-Interim produce overestimated LH fluxes. The poor diurnal cycle in wind speed in CFSR is also reflected in the poor diurnal cycle in its LH flux, with a maximum at 1930 UTC as opposed to the 0430 UTC maximum observed (Fig. 10e). SH flux in ERA-Interim is way too high because of its lower 2-m temperatures. The reanalysis SH fluxes steadily increase

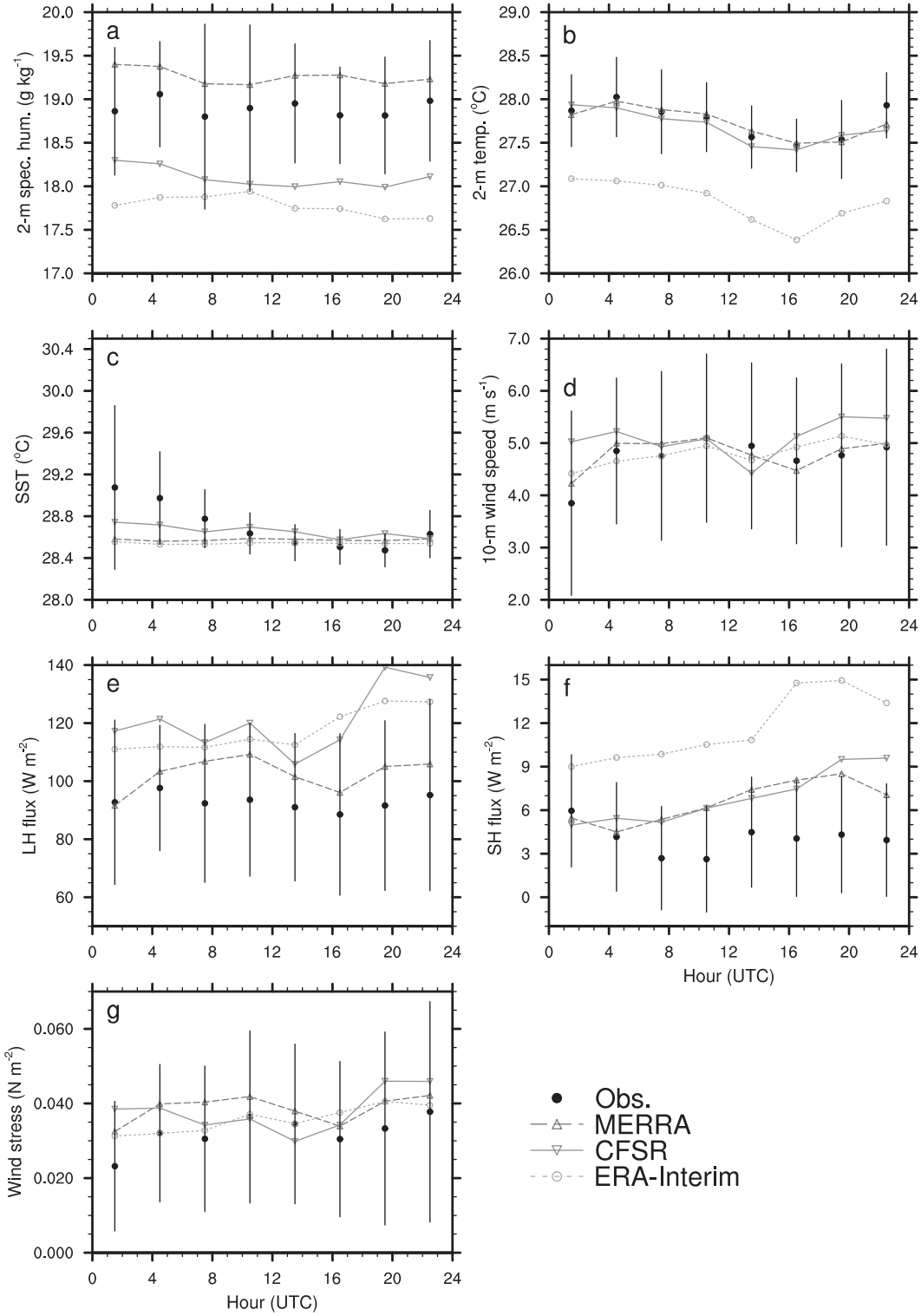

FIG. 10. The 3-h mean (a) 2-m specific humidity, (b) 2-m air temperature, (c) SST, (d) 10-m wind speed, (e) LH flux, (f) SH flux, and (g) wind stress from observations (with ±1 standard deviation represented by the vertical lines), MERRA, CFSR, and ERA-Interim over the course of KWAJEX.

over the course of the day because of increasing air–sea temperature differences due to the small diurnal cycle in SST (Fig. 10f). The diurnal cycle of wind stress closely follows that of the 10-m wind speed, and the three reanalyses overestimate wind stress most of the time (Fig. 10g).

6. Discussion and conclusions

Intercomparison of the ocean surface turbulent fluxes in six reanalyses (MERRA, ERA-40, ERA-Interim, NCEP–NCAR, NCEP–DOE, and CFSR), four satellite-derived products (GSSTF2, GSSTF2b, J-OFURO, and HOAPS), and one combined product (OAFlux) reveals that the product flux biases are a combination of two uncertainties: bulk variable caused and residual [Eq. (4)]. The residual uncertainties generally dominate for all-cruise LH flux biases, while the bulk-variable-caused uncertainties tend to dominate in most of the all-cruise SH flux and wind stress biases (Table 1).

The bulk-variable-caused uncertainties in fluxes are a combined effect of the errors in the bulk variables used in the products. While 2-m specific humidity errors over all the cruises are small because of equal and opposite regional errors (Fig. 8a), wind speed errors are quite strong from NCEP–DOE, CFSR, and the satellite-derived products (Fig. 8b). SST and 2-m air temperature biases are generally low except for a few cruises, particularly in 2-m temperature during SCOPE for all of the reanalyses that have a warming during the latter half of the cruise, which is not observed (not shown).

For the satellite-derived products, these bulk variable errors are simply produced by inaccuracies in the retrieval. In the reanalyses, there are also measurement errors and model errors that include uncertainties from the physical parameterizations other than the surface flux algorithm and errors in the assimilation of data. These can result in an overall bias, as in the specific humidity underestimations by CFSR and ERA-Interim or air temperature underestimation by ERA-Interim during KWAJEX, or in temporally dependent biases, as in the wind speed biases in CFSR and ERA-Interim during KWAJEX (Fig. 10). The latter can be partially caused by shocks added to the model when the assimilation is performed. A comparison of the hourly bulk variables from MERRA and CFSR during KWAJEX shows that there are periodically unrealistic jumps in the 10-m wind speed at some of the assimilation times in CFSR, whereas there are none in MERRA (not shown). This suggests the need to minimize shocks to the model at assimilation times, as is done in MERRA, through the use of the incremental analysis update (IAU) technique.

The residual uncertainties include contributions from the algorithm uncertainty (i.e., the uncertainty as to the accuracy of the bulk algorithm used to calculate the flux) and measurement uncertainty. Still, a comparison of the residual uncertainties between products can reveal some valuable information about the algorithm uncertainties. For instance, the reduction in the residual uncertainty in all the fluxes from NCEP–NCAR and NCEP–DOE to CFSR shows that the inclusion of the Zeng et al. (1998) heat and moisture roughness lengths has reduced the algorithm uncertainties in the NCEP flux algorithm (Table 1). The residual uncertainties from the satellite-derived and combined products are also lower than that of the ECMWF and NCEP reanalyses for LH flux and that of the NCEP reanalyses for SH flux. Several of these products (J-OFURO, HOAPS, and OAFlux) utilize the COARE 3.0 algorithm, which was found to be one of the least problematic overall (Brunke et al. 2003).

Finally, the ranking of these widely used product fluxes according to Eq. (5) provides an assessment of the best-performing products (Table 2). MERRA is in category A for all three fluxes. Also in category A are ERA-Interim for LH flux and wind stress, GSSTF2b for LH and SH fluxes, ERA-40 for LH flux, OAFlux for SH flux, and NCEP–DOE for wind stress.

These products qualify because they have some of the lowest biases and/or SDEs. For some of these products, their good performance is due to having very low biases that have nearly equal contributions from bulk variable and residual uncertainties. For instance, this is true for all-cruise LH and SH fluxes in GSSTF2b as well as for all-cruise SH fluxes in MERRA (Table 1). Plus, there are still large bulk-variable-caused uncertainties in some of the products that generally increase in high wind speed conditions (Table 1; Fig. 6). Thus, significant improvement is still needed in both algorithms and the retrieval of bulk variables to reduce the total uncertainty in product LH and SH fluxes to within 5 W m^{-2}.

In general, the new generation of flux datasets is as good or better than the older generations of products. For instance, the new version of GSSTF (i.e., GSSTF2b) represents a significant improvement over GSSTF2 (e.g., Table 1). In contrast, ERA-40 already performed reasonably well, so there is not much improvement in ERA-Interim (e.g., Table 1). The results from CFSR, which is the only reanalysis involving a coupled atmosphere–ocean data assimilation, are not that bad considering that it falls into category B for all three fluxes (Table 1). However, further work may still be needed to realize the full potential of ocean–atmosphere coupled reanalysis.

While the current generation of satellite-derived and combined products (e.g., GSSTF2b and OAFlux) are generally better than the earlier generation of reanalyses

TABLE 4. The mean all-cruise difference in surface downward LW and SW radiative fluxes between the reanalyses and ship observations.

Product	LW flux	SW flux
MERRA	−9.5	5.8
ERA-40	6.2	12.1
ERA-Interim	−1.2	21.8
NCEP–NCAR	3.5	38.3
NCEP–DOE	6.8	15.2
CFSR	0.1	30.1

(NCEP–NCAR and NCEP–DOE), they are comparable or slightly worse than the new generation of reanalysis (particularly MERRA). This is because 1) the data assimilation method has been improved from the first generation to the current generation of reanalyses and 2) more satellite data (including all of those used in the satellite-derived products plus others that were not used) are assimilated in the new generation of reanalyses. Furthermore, when compared with satellite-derived and combined products, the new reanalyses provide higher temporal resolution (3 hourly in ERA-Interim and hourly in MERRA and CFSR) than the previous generation of reanalyses (6 hourly) or the satellite-derived and combined products (daily or 12 hourly).

Another shortcoming of the satellite-derived products is that they do not generally provide radiative fluxes, which are also needed to study the surface energy budget or to force an OGCM, whereas they are generally not provided in the satellite-derived products [HOAPS does provide the net longwave (LW) flux but not the net shortwave (SW) flux]. Table 4 presents the mean difference in surface downward LW and SW radiative fluxes between the reanalysis values and ship observations from all of the cruises. LW radiation is very well constrained in the reanalyses with CFSR having the lowest bias, but a few reanalyses have high biases ($>$20 W m^{-2}) in SW radiation. In fact, CFSR has the highest bias of all of the reanalyses for SW radiation. In contrast, if the satellite-derived fluxes are desired, then one could use them with another satellite radiative flux dataset, such as the International Satellite Cloud Climatology Project (ISCCP; Zhang et al. 2004), as is also included with the OAFlux data.

Another problem facing reanalyses is inconsistencies that result with changes with the satellite data ingested as newer and improved satellites come online. Such ''jumps'' can be seen in MERRA, for instance, in Robertson et al. (2011). For example, the moisture corrections change suddenly from drying to moistening over the course of the MERRA assimilation, with most of that change happening in late 1998, when the assimilation of *NOAA-15* AMSU-A data began and to a lesser extent in late 1987 with the incorporation of the SSM/I data. Bosilovich et al. (2011) report that changes in evaporation due to the change in the analysis moisture increments with the ingestion of AMSU-A as compared with before depend upon the region. Such changes result in LH flux biases that are more positive for all of the 1999 cruises except for PACS Flux '99, which is negative (not shown). PACS Flux '99 corresponds to an area of decreased evaporation after late 1998, and the other cruises correspond to areas of weak or positive change in evaporation afterward (see Bosilovich et al. 2011, their Fig. 12b). Another possible concern is the increasing number of data from multiple SSM/I satellites ingested, which would affect near-surface wind speed. There is no effective difference in the wind speed biases and SDEs between the cruises except by region (Figs. 8b, 9b), but there were at least two SSM/I satellites ingested into MERRA during all of these cruises (http://gmao.gsfc.nasa.gov/research/merra/IMAGES/MERRA_Satellite_data_streams.jpg). Such biases may be more substantial if any cruises in 1987–90, when there was only one SSM/I satellite, or even pre-SSM/I (before 1987) cruises were used.

Despite these concerns, the current generation of reanalyses (MERRA, ERA-Interim, and CFSR) still holds promise for the future direction in the development of ocean surface turbulent flux datasets because of their high frequency and all of the satellite data that they ingest. Such datasets would be derived from a combination of the high-frequency bulk variables from the new generation of reanalysis products (MERRA, ERA-Interim, and CFSR) and computing high-frequency fluxes using one of the least problematic bulk algorithms discussed in Brunke et al. (2003).

Acknowledgments. This work was supported by NASA under Grant NNX09A021G. Dr. C. Fairall is thanked for providing the ship cruise data and the COARE 3.0 algorithm used here. We also thank Drs. A. Beljaars, H.-L. Pan, and S.-H. Chou for providing their algorithms. Dr. A. Beljaars and two anonymous reviewers are also thanked for their helpful comments. Finally, NCAR is thanked for providing the computing resources to download the reanalysis data used here.

APPENDIX

The Product Bulk Algorithms

The bulk algorithms use Eqs. (1)–(3) to calculate the turbulent surface fluxes compared here. The exchange coefficients (i.e., C_H, C_E, and C_D) are defined as

$$C_H = \frac{k^2}{\left[\ln\left(\frac{z}{z_{\mathrm{om}}}\right) - \psi_m\left(\frac{z}{L}\right)\right]\left[\ln\left(\frac{z}{z_{\mathrm{ot}}}\right) - \psi_h\left(\frac{z}{L}\right)\right]}, \quad \text{(A1)}$$

$$C_E = \frac{k^2}{\left[\ln\left(\frac{z}{z_{\mathrm{om}}}\right) - \psi_m\left(\frac{z}{L}\right)\right]\left[\ln\left(\frac{z}{z_{\mathrm{oq}}}\right) - \psi_h\left(\frac{z}{L}\right)\right]}, \quad \text{(A2)}$$

and

$$C_H = \frac{k^2}{\left[\ln\left(\frac{z}{z_{\mathrm{om}}}\right) - \psi_m\left(\frac{z}{L}\right)\right]^2}, \quad \text{(A3)}$$

where k is the von Kármán constant of 0.4; z is the height of the lowest model level above the ground; z_{om}, z_{ot}, and z_{oq} are the roughness lengths for momentum, temperature, and humidity, respectively; L is the Obukhov length; and ψ_m and ψ_h are stability functions. It is the parameterization of these coefficients that causes most of the uncertainty between algorithms, although there are differences in how other parameters are set that could be substantial in certain circumstances (Zeng et al. 1998; Brunke et al. 2002, 2003). Here, we focus on the differences in the formulation of the exchange coefficients. Other details of the algorithms are explained in the three studies mentioned above.

The algorithm in the GEOS-5 AGCM, which is used to produce MERRA, as described in Helfand and Schubert (1995), uses stability functions from Clarke (1970) for stable conditions and from Panofsky and Dutton (1984) for unstable conditions. The roughness length for momentum is a function of the friction velocity u_* as follows:

$$z_{\mathrm{om}} = \frac{a_1}{u_*} + a_2 + a_3 u_* + a_4 u_*^2 + a_5 u_*^3, \quad \text{(A4)}$$

where the coefficients a_1, a_2, a_3, a_4, and a_5 are taken from Large and Pond (1981) for moderate to large wind speeds, Kondo (1975) for weak wind speeds, and interpolated between the two for wind speeds in between. The roughness lengths for temperature and humidity are

$$\ln\left(\frac{z_{\mathrm{om}}}{z_{\mathrm{ot}}}\right) = \ln\left(\frac{z_{\mathrm{om}}}{z_{\mathrm{oq}}}\right) = 0.72(\mathrm{Re}_* - 0.135)^{1/2}, \quad \text{(A5)}$$

where Re_* is the roughness Reynolds number calculated from the friction velocity, roughness length for momentum, and kinematic viscosity of air ν as $u_* z_{\mathrm{om}}/\nu$.

In the algorithm in the ECMWF operational model (Beljaars 1995) used to produce both ERA-40 and ERA-Interim, the stability functions are taken from Holtslag and de Bruin (1988) for stable conditions and as used in Dyer (1974) for unstable conditions. The roughness lengths are

$$z_{\mathrm{om}} = 0.018\frac{u_*^2}{g} + \frac{1.65 \times 10^{-6}}{u_*}, \quad \text{(A6)}$$

$$z_{\mathrm{ot}} = \frac{6 \times 10^{-6}}{u_*}, \quad \text{and} \quad \text{(A7)}$$

$$z_{\mathrm{oq}} = \frac{9.3 \times 10^{-6}}{u_*}. \quad \text{(A8)}$$

Note that z_{om} is actually computed from an ocean wave model (ECMWF 2007) in the ECMWF operational model and reanalysis, but (A6) is used in our uncertainty decomposition analysis in section 4.

The algorithm in the NCEP operational model used to produce the NCEP–NCAR and NCEP–DOE reanalyses extends Monin–Obukhov similarity to include empirical profile equations for various stability regimes. The roughness lengths are parameterized as follows:

$$z_{\mathrm{om}} = 0.014\frac{u_*^2}{g} \quad \text{and} \quad \text{(A9)}$$

$$\ln\left(\frac{z_{\mathrm{om}}}{z_{\mathrm{ot}}}\right) = \ln\left(\frac{z_{\mathrm{om}}}{z_{\mathrm{oq}}}\right) = \frac{-1.076 + 0.7045\ln(\mathrm{Re}_*) - 0.058\,08[\ln(\mathrm{Re}_*)]^2}{1 - 0.1954\ln(\mathrm{Re}_*) + 0.009\,999[\ln(\mathrm{Re}_*)]^2}. \quad \text{(A10)}$$

The algorithm in the CFS used to produce the CFSR is the same as the previous one used for NCEP–NCAR and NCEP–DOE, except that the Zeng et al. (1998) roughness lengths for heat and moisture are used instead as shown:

$$\ln\left(\frac{z_{\mathrm{om}}}{z_{\mathrm{ot}}}\right) = \ln\left(\frac{z_{\mathrm{om}}}{z_{\mathrm{oq}}}\right) = 2.67\,\mathrm{Re}_*^{1/4} - 2.57. \qquad \text{(A11)}$$

The bulk algorithm used to produce both versions of GSSTF here was based on Chou (1993) with the addition of the salinity effect to surface saturated humidity and a change to the parameters used for the von Kármán constant (Chou et al. 2003). The stability functions are taken from Large and Pond (1981) for stable conditions and Businger et al. (1971) for unstable conditions. The roughness length for momentum is

$$z_{\mathrm{om}} = \frac{0.0144u_*^2}{g} + \frac{0.11\nu}{u_*}, \qquad \text{(A12)}$$

while the following roughness lengths for heat and moisture are from Liu et al. (1979):

$$\frac{z_{\mathrm{ot}}u_*}{\nu} = a_1\mathrm{Re}_*^{b_1} \quad \text{and} \qquad \text{(A13)}$$

$$\frac{z_{\mathrm{oq}}u_*}{\nu} = a_2\mathrm{Re}_*^{b_2}. \qquad \text{(A14)}$$

The coefficients a_1, a_2, b_1, and b_2 are taken from Table 1 in Liu et al. (1979).

J-OFURO, HOAPS, and OAFlux utilize the COARE 3.0 algorithm (Fairall et al. 1996, 2003), which was found to be one of the least problematic in Brunke et al. (2003). For unstable conditions, the Grachev et al. (2000) stability functions are used, whereas under stable conditions, the stability functions are taken from Beljaars and Holtslag (1991). The roughness length for momentum is parameterized as

$$z_{\mathrm{om}} = \frac{au_*^2}{g} + \frac{0.11\nu}{u_*}, \qquad \text{(A15)}$$

where a varies with wind speed, such that

$$a = \begin{cases} 0.011 & \text{for} \quad U \le 10\,\mathrm{m\,s^{-1}} \\ 0.011 + \dfrac{0.007}{8}(U - 10) & \text{for} \quad 10 < U < 18\,\mathrm{m\,s^{-1}} \\ 0.018 & \text{for} \quad U \ge 18\,\mathrm{m\,s^{-1}} \end{cases}. \qquad \text{(A16)}$$

The roughness lengths for heat and momentum are equal to the smaller of $5.5 \times 10^{-5}\,\mathrm{Re}_*^{-0.6}$ or 1.1×10^{-5} m.

REFERENCES

Alpert, J. C., M. Kanamitsu, P. M. Caplan, J. G. Sela, G. H. White, and E. Kalnay, 1988: Mountain induced gravity wave drag parameterization in the NMC medium-range forecast model. Preprints, *Eighth Conf. on Numerical Weather Prediction,* Baltimore, MD, Amer. Meteor. Soc., 726–733.

——, S.-Y. Hong, and Y.-J. Kim, 1996: Sensitivity of cyclogenesis to lower tropospheric enhancement of gravity wave drag using the environmental modeling center medium range model. Preprints, *11th Conf. on Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., 322–323.

Andersson, A., K. Fennig, C. Klepp, S. Bakan, H. Graßl, and J. Schulz, 2010: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data—HOAPS-3. *Earth Syst. Sci. Data,* **2,** 215–234.

Beljaars, A. C. M., 1995: The parameterization of surface fluxes in large scale models under free convection. *Quart. J. Roy. Meteor. Soc.,* **121,** 255–270.

——, and A. A. M. Holtslag, 1991: Flux parameterization over land surfaces for atmospheric models. *J. Appl. Meteor.,* **30,** 327–341.

Bentamy, A., K. B. Katsaros, A. M. Mestas-Nuñez, W. M. Drennan, E. B. Forder, and H. Roquet, 2003: Satellite estimates of wind speed and latent heat flux over the global oceans. *J. Climate,* **16,** 637–656.

Berry, D. I., and E. C. Kent, 2009: A new air–sea interaction gridded dataset from ICOADS with uncertainty estimates. *Bull. Amer. Meteor. Soc.,* **90,** 645–656.

Bloom, S., L. Takacs, A. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.,* **124,** 1256–1271.

Bosilovich, M. G., F. R. Robertson, and J. Chen, 2011: Global energy and water budgets in MERRA. *J. Climate,* in press.

Bourras, D., 2006: Comparison of five satellite-derived latent heat flux products to moored buoy data. *J. Climate,* **19,** 6291–6313.

——, L. Eymard, and W. T. Liu, 2002: A neural network to estimate the latent heat flux over oceans from satellite observations. *Int. J. Remote Sens.,* **23,** 2405–2423.

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

——, C. W. Fairall, X. Zeng, L. Eymard, and J. A. Curry, 2003: Which bulk aerodynamic algorithms are least problematic in computing ocean surface turbulent fluxes? *J. Climate,* **16,** 619–635.

——, X. Zeng, V. Misra, and A. Beljaars, 2008: Integration of a prognostic sea surface skin temperature scheme into weather and climate models. *J. Geophys. Res.,* **113,** D21117, doi:10.1029/2008JD010607.

Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley, 1971: Flux profile relationships in the atmospheric surface layer. *J. Atmos. Sci.,* **28,** 181–189.

Casey, K. S., T. B. Brandon, P. Cornillon, and R. Evans, 2010: The past, present, and future of the AVHRR Pathfinder SST program. *Oceanography from Space: Revisited,* V. Barale, J. F. R. Gower, and L. Alberotanza, Eds., Springer, 323–341.

Chou, S.-H., 1993: A comparison of airborne eddy correlation and bulk aerodynamic methods for ocean-air turbulent fluxes during cold-air outbreaks. *Bound.-Layer Meteor.,* **64,** 75–100.

——, R. M. Atlas, C.-L. Shie, and J. Ardizzone, 1995: Estimates of surface humidity and latent heat fluxes over oceans from SSM/I data. *Mon. Wea. Rev.,* **123,** 2405–2425.

——, C.-L. Shie, R. M. Atlas, and J. Ardizzone, 1997: Air–sea fluxes retrieved from Special Sensor Microwave Imager data. *J. Geophys. Res.,* **102,** 12 705–12 726.

——, E. Nelkin, J. Ardizzone, R. M. Atlas, and C.-L. Shie, 2003: Surface turbulent heat and momentum fluxes over global

oceans based on the Goddard satellite retrievals, version 2 (GSSTF2). *J. Climate,* **16,** 3256–3273.

Clarke, R. H., 1970: Observational studies in the atmospheric boundary layer. *Quart. J. Roy. Meteor. Soc.,* **96,** 91–114.

Curry, J. A., and Coauthors, 2004: SEAFLUX. *Bull. Amer. Meteor. Soc.,* **85,** 409–424.

da Silva, A., C. C. Young, and S. Levitus, 1994: *Algorithms and Procedures*. Vol. 1, *Atlas of Surface Marine Data 1994,* NOAA Atlas NESDIS 6, 83 pp.

Dyer, A. J., 1974: A review of flux-profile relationships. *Bound.-Layer Meteor.,* **7,** 363–372.

ECMWF, 2007: Part VII: ECMWF wave model. IFS Documentation Cy31r1, 56 pp. [Available online at http://www.ecmwf.int/research/ifsdocs/CY31r1/WAVES/IFSPart7.pdf.]

Fairall, C. W., E. F. Bradley, D. P. Rogers, J. B. Edson, and G. S. Young, 1996: Bulk parameterization of air-sea fluxes for Tropical Ocean-Global Atmosphere Coupled-Ocean Atmosphere Response Experiment. *J. Geophys. Res.,* **101,** 3747–3764.

——, ——, 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.

Grachev, A. A., C. W. Fairall, and E. F. Bradley, 2000: Convective profile constants revisited. *Bound.-Layer Meteor.,* **94,** 495–515.

Helfand, H. M., and S. D. Schubert, 1995: Climatology of the simulated Great Plains low-level jets and its contribution to the continental moisture budget of the United States. *J. Climate,* **8,** 784–806.

Holtslag, A. A. M., and H. A. R. de Bruin, 1988: Applied modeling of the nighttime surface energy balance over land. *J. Appl. Meteor.,* **27,** 689–704.

Hong, S.-Y., and H.-L. Pan, 1998: Convective trigger function for a mass-flux cumulus parameterization scheme. *Mon. Wea. Rev.,* **126,** 2599–2620.

Jackson, D. L., and G. A. Wick, 2010: Near-surface air temperature retrieval derived from AMSU-A and sea surface temperature observations. *J. Atmos. Oceanic Technol.,* **27,** 1769–1776.

——, ——, and J. J. Bates, 2006: Near-surface retrieval of air temperature and specific humidity using multisensory microwave satellite observations. *J. Geophys. Res.,* **111,** D10306, doi:10.1029/2005JD006431.

Josey, S. A., E. C. Kent, and P. K. Taylor, 1999: New insights into the ocean heat budget closure problem from analysis of the SOC air–sea flux climatology. *J. Climate,* **12,** 2856–2880.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

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.

Kim, Y.-J., and A. Arakawa, 1995: Improvement of orographic gravity wave parameterization using a mesoscale gravity wave model. *J. Atmos. Sci.,* **52,** 1875–1902.

Kondo, J., 1975: Air-sea bulk transfer coefficients in diabatic conditions. *Bound.-Layer Meteor.,* **9,** 91–112.

Krasnopolsky, V. M., L. C. Breaker, and W. H. Gemmil, 1995: A neural network as a nonlinear transfer function model for retrieving surface wind speeds from the Special Sensor Microwave Imager. *J. Geophys. Res.,* **100,** 11 033–11 045.

Kubota, M., N. Iwasaka, S. Kizu, M. Konda, and K. Kutsuwada, 2002: Japanese Ocean Flux Data Sets with Use of Remote Sensing Observations (J-OFURO). *J. Oceanogr.,* **58,** 213–225.

——, A. Kano, H. Muramatsu, and H. Tomita, 2003: Intercomparison of various surface latent heat flux fields. *J. Climate,* **16,** 670–678.

Large, W. G., and S. Pond, 1981: Open ocean momentum flux measurements in moderate to strong winds. *J. Phys. Oceanogr.,* **11,** 324–336.

Liu, W. T., K. B. Katsaros, and J. A. Businger, 1979: Bulk parameterization of air-sea exchanges of heat and water vapor including the molecular constraints at the interface. *J. Atmos. Sci.,* **36,** 1722–1735.

——, A. Zhang, and J. Bishop, 1994: Evaporation and solar irradiance as regulators of sea surface temperature in annual and interannual changes. *J. Geophys. Res.,* **99,** 12 623–12 638.

Moorthi, S., H. L. Pan, and P. Caplan, 2001: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. NWS Tech. Procedures Bull. 484, 14 pp. [Available online at http://www.nws.noaa.gov/om/tpb/484.pdf.]

Murray, F. W., 1967: On the computation of saturation vapor pressure. *J. Appl. Meteor.,* **6,** 203–204.

Pan, H.-L., and W.-S. Wu, 1995: Implementing a mass flux convective parameterization package for the NMC medium-range forecast model. NMC Office Note 409, 40 pp. [Available online at http://www.emc.ncep.noaa.gov/officenotes/FullTOC.html#1990.]

Panofsky, H. A., and J. A. Dutton, 1984: *Atmospheric Turbulence: Models and Methods for Engineering Applications*. John Wiley and Sons, 397 pp.

Pedreros, R., G. Dardier, H. Dupuis, H. C. Graber, W. M. Drennan, A. Weill, C. Guérin, and P. Nacass, 2003: Momentum and heat fluxes via the eddy correlation method on the R/V *L'Atalante* and an ASIS buoy. *J. Geophys. Res.,* **108,** 3339, doi:10.1029/2002JC001449.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Powell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.,* **108,** 4407, doi:10.1029/2002JD002670.

——, ——, ——, ——, ——, D. P. Rowell, E. C. Kent, and A. Kaplan, 2006: Met Office - GISST/MOHMATN4/MOHSST6 - Global ice coverage and SST (1856-2006). [Available online at http://badc.nerc.ac.uk/data/gisst.]

Renfrew, I. A., G. W. K. Moore, P. S. Guest, and K. Bumke, 2002: A comparison of surface layer and surface turbulent flux observations over the Labrador Sea with ECMWF analyses and NCEP reanalyses. *J. Phys. Oceanogr.,* **32,** 383–400.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses. *J. Climate,* **7,** 929–948.

——, ——, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperature. *J. Climate,* **20,** 5473–5496.

Rienecker, M. R., and Coauthors, 2007: The GEOS-5 Data Assimilation System - Documentation on versions 5.0.1, 5.1.0, and 5.2.0. NASA Tech Rep. NASA/TM-2007-104606, 95 pp.

——, and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate,* **24,** 3624–3648.

Roberts, J. B., C. A. Clayson, F. R. Robertson, and D. L. Jackson, 2010: Predicting near-surface atmospheric variables from Special Sensor Microwave/Imager using neural networks with a first-guess approach. *J. Geophys. Res.,* **115,** D19113, doi:10.1029/2009JD013099.

Robertson, F. R., M. G. Bosilovich, J. Chen, and T. L. Miller, 2011: The effect of satellite observing system changes on MERRA water and energy fluxes. *J. Climate,* **24,** 5197–5217.

Rosenfield, J. E., M. R. Schoeberl, and M. A. Geller, 1987: A computation of the stratospheric diabatic circulation using an accurate radiative transfer model. *J. Atmos. Sci.,* **44,** 859–876.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System reanalysis. *Bull. Amer. Meteor. Soc.,* **91,** 1015–1057.

Schlüssel, P., L. Schanz, and G. Englisch, 1995: Retrieval of latent heat flux and longwave irradiance at the sea surface from SSM/I and AVHRR measurements. *Adv. Space Res.,* **16,** 107–116.

Schulz, J., P. Schlüssel, and H. Grassl, 1993: Water vapor in the atmospheric boundary layer over oceans from SSM/I measurements. *Int. J. Remote Sens.,* **14,** 2773–2789.

Shie, C.-L., 2010: Science background for the reprocessing and Goddard Satellite-Based Surface Turbulent Fluxes (GSSTF2b) data set for global water and energy cycle research. Goddard Earth Sciences Data and Information Services Center, 18 pp. [Available online at http://disc.sci.gsfc.nasa.gov/measures/documentation/Science-of-the-data.pdf.]

——, and K. Hilburn, 2011: A newly revised satellite-based global air–sea surface turbulent fluxes dataset and its dependence on the SSM/I brightness temperature. *Proc. 2011 IEEE IGARSS,* Vancouver, BC, Canada, IEEE, 4 pp.

——, and Coauthors, 2009: A note on reviving the Goddard Satellite-Based Surface Turbulent Fluxes (GSSTF) dataset. *Adv. Atmos. Sci.,* **26,** 1071–1080.

Simmons, A., S. Uppala, D. Dee, and S. Kobayashi, 2006: ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter,* No. 110, ECMWF, Reading, United Kingdom, 25–35.

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.

Suarez, M., and Coauthors, 2008: File specification for MERRA products, version 2.1. NASA GMAO Rep., 96 pp. [Available online at http://gmao.gsfc.nasa.gov/research/merra/MERRA_FileSpec_DRAFT_09_02_2008.pdf.]

Sundqvist, H., E. Berge, and J. E. Kristjansson, 1989: Condensation and cloud studies with mesoscale numerical weather prediction model. *Mon. Wea. Rev.,* **117,** 1641–1657.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012.

Wang, W., and M. J. McPhaden, 2001: What is the mean seasonal cycle of surface heat flux in the equatorial Pacific? *J. Geophys. Res.,* **106,** 837–857.

Weller, R. A., and S. P. Anderson, 1996: Surface meteorology and air–sea fluxes in the western equatorial Pacific warm pool during the TOGA Coupled Ocean–Atmosphere Response Experiment. *J. Climate,* **9,** 1959–1990.

Wells, N., and S. King-Hele, 1990: Parameterization of tropical ocean heat flux. *Quart. J. Roy. Meteor. Soc.,* **116,** 1213–1224.

Wentz, F. J., 1997: A well calibrated ocean algorithm for SSM/I. *J. Geophys. Res.,* **102,** 8703–8718.

——, L. Ricciardulli, K. Hilburn, and C. Mears, 2007: How much more rain will global warming bring? *Science,* **317,** 233–235.

Worley, S. J., S. D. Woodruff, R. W. Reynolds, S. J. Lubker, and N. Lott, 2005: ICOADS release 2.1 data and products. *Int. J. Climatol.,* **25,** 823–842.

Yu, L., and R. A. Weller, 2007: Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005). *Bull. Amer. Meteor. Soc.,* **88,** 527–539.

——, X. Jin, and R. A. Weller, 2008: Multidecade global flux datasets from the objectively analyzed air-sea fluxes (OAFlux) project: Latent and sensible heat fluxes, ocean evaporation, and related surface meteorological variables. OAFlux Project Tech. Rep. OA-2008-01, 64 pp.

Zeng, X., M. Zhao, and R. E. Dickinson, 1998: Intercomparison of bulk aerodynamic algorithms for the computation of sea surface fluxes using TOGA COARE and TAO data. *J. Climate,* **11,** 2628–2644.

Zhang, Y., W. B. Rossow, A. A. Lacis, V. Oinas, and M. I. Mishchenko, 2004: Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *J. Geophys. Res.,* **109,** D19105, doi:10.1029/2003JD004457.

Zhao, Q., and F. H. Carr, 1997: A prognostic cloud scheme for operational NWP models. *Mon. Wea. Rev.,* **125,** 1931–1953.