Mean Biases, Variability, and Trends in Air–Sea Fluxes and Sea Surface Temperature in the CCSM4

SUSAN C. BATES,* BAYLOR FOX-KEMPER,+ STEVEN R. JAYNE,# WILLIAM G. LARGE,* SAMANTHA STEVENSON,@ AND STEPHEN G. YEAGER*

* *National Center for Atmospheric Research,& Boulder, Colorado*
+ *Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado*
# *Woods Hole Oceanographic Institution, Woods Hole, Massachusetts*
@ *International Pacific Research Center, University of Hawaii at Manoa, Honolulu, Hawaii*

(Manuscript received 10 August 2011, in final form 26 May 2012)

ABSTRACT

Air–sea fluxes from the Community Climate System Model version 4 (CCSM4) are compared with the Coordinated Ocean-Ice Reference Experiment (CORE) dataset to assess present-day mean biases, variability errors, and late twentieth-century trend differences. CCSM4 is improved over the previous version, CCSM3, in both air–sea heat and freshwater fluxes in some regions; however, a large increase in net shortwave radiation into the ocean may contribute to an enhanced hydrological cycle. The authors provide a new baseline for assessment of flux variance at annual and interannual frequency bands in future model versions and contribute a new metric for assessing the coupling between the atmospheric and oceanic planetary boundary layer (PBL) schemes of any climate model. Maps of the ratio of CCSM4 variance to CORE reveal that variance on annual time scales has larger error than on interannual time scales and that different processes cause errors in mean, annual, and interannual frequency bands. Air temperature and specific humidity in the CCSM4 atmospheric boundary layer (ABL) follow the sea surface conditions much more closely than is found in CORE. Sensible and latent heat fluxes are less of a negative feedback to sea surface temperature warming in the CCSM4 than in the CORE data with the model's PBL allowing for more heating of the ocean's surface.

1. Introduction

The coupling between the atmosphere and ocean is a major player in the earth's climate system and governor of climate change. The former has a limited capacity to store water and heat but is connected to the ocean, which is effectively an infinite reservoir of water and has more heat capacity in only its upper few meters than exists in the entire atmosphere. The direct coupling of the planetary boundary layers (PBLs) is accomplished through the air–sea fluxes. In nature, the global, net air–sea fluxes of heat and freshwater are near zero (Large and Yeager 2009) and much smaller than the uncertainty in observational estimates. However, significant climate trends can be sustained by small imbalances in the air–sea heat fluxes and the freshwater fluxes into the ocean. The evaporation of water from the ocean determines the strength of the earth's hydrological cycle, and the climate effects of imbalance with the surface water fluxes of precipitation and runoff are changes in ocean salinity and thus density and stratification.

Increases in atmospheric greenhouse gas (GHG) concentrations, stratospheric aerosols from a volcanic eruption, and changes in the earth's orbit all perturb the earth's radiative heat balance. For example, GHG increases cause no change in the incoming solar, but the greater absorption of outgoing longwave radiation produces tropospheric warming, leading to an increase in the downward longwave radiation. Over the ocean, any of this longwave heat flux reaching the surface would increase sea surface temperature (SST) until the SST-dependent cooling increases enough to reestablish

& The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Susan C. Bates National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.
E-mail: bates@ucar.edu

DOI: 10.1175/JCLI-D-11-00442.1

balance. This occurs through the surface fluxes that depend on PBL properties, namely, latent and sensible heat and surface upward longwave emission.

Coupled climate models such as the Community Climate System Model (CCSM) strive to achieve an accurate representation of atmosphere–ocean coupling and its variation with climate change. Model confidence would increase if it could be shown that agreement with SST was achieved with realistic air–sea fluxes. However, attempts have run into the problem of uncertain (Taylor 2000) and scattered flux estimates, as illustrated by Roske's (2006) valuable compilation of available air–sea flux datasets. For example, Wittenberg et al. (2006) find the range in estimates of the heat flux across the equatorial Pacific to be too large to assess model fidelity. Nonetheless, an examination of the realism of air–sea flux fields should become a necessary part of any comprehensive assessment of coupled model skill, because they reflect coupling mechanisms.

Our approach to the problems with observed fluxes is to compare model fields with the best available component flux data obtained from multiple sources as advocated by Gleckler et al. (2008). The flux dataset computed from the forcing developed for the Coordinated Ocean-Ice Reference Experiments (CORE) combined with the Hurrell et al. (2008) SST product is used. This CORE (version 2) dataset, is described and evaluated by Large and Yeager (2009). The CORE forcing is based largely on the vector wind, air temperature, and humidity from the National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996), but it uses data from alternate sources for radiation, precipitation, and sea ice concentration as well as for the critical step of correcting mean biases. It has been used to force a number of different coupled ocean–sea ice models and the solutions are compared by Griffies et al. (2009). The CORE fluxes are well suited for our purposes of evaluating coupled climate model mean fluxes and aspects of the variability because they satisfy the following requirements: they include global estimates of momentum, heat, and freshwater fluxes and their components; the net global heat and water fluxes are near zero, consistent with observations (Large and Yeager 2009); the time series span multiple decades; fluxes are available at high frequency; and implied ocean heat transports are consistent with estimates based on ocean measurements.

In this paper, we use the CORE heat and freshwater fluxes to evaluate the mean state, variability, and trends of the air–sea fluxes in an ensemble of twentieth-century (20C) simulations of the CCSM version 4 (CCSM4) in response to specified GHG emissions. Mean air–sea heat and freshwater flux fields and the ensemble spread are examined, so they can be related to the previous evaluation of CCSM version 3 (CCSM3; Large and Danabasoglu 2006) and its conclusion that errors in these fields were largely responsible for sea surface temperature and salinity biases, respectively. This attribution was based on numerical experiments involving select substitution of model fluxes and related parameters with observations. Stine et al. (2009) point out the poor quality of the annual cycle in Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) GCMs, and improving interannual variability has also received recent attention in CCSM (Neale et al. 2008; Stevenson et al. 2010, 2012). We examine flux variability in CCSM4 on these time scales, a first for CCSM because having the CORE fluxes only now makes this comparison meaningful. Since there are no comparable CCSM3 analyses, the results will serve as a benchmark for future model versions. Although no comparisons of variability with CCSM3 are made within this text, Deser et al. (2012) examine the interannual time scale in the context of the representation of ENSO in CCSM4 along with some comparisons with earlier CCSM versions. Finally, the trends in SST in the 20C simulations and the associated behavior of the heat flux and its components are compared to the negative flux feedbacks in the CORE fluxes in response to the increasing trends in SST observed since the 1980s (Large and Yeager 2012). There are no comparisons with CCSM3 trends because the different spinup procedures used in CCSM3 and CCSM4 (Gent et al. 2011) significantly impact model trends (Danabasoglu et al. 2011).

2. Data and analysis techniques

Most of the historical CORE air–sea flux data are available globally, every 6 h from 1948 through 2007. The notable exceptions are precipitation (monthly from 1979), downward longwave and solar radiation (daily from July 1983), and sea ice concentration (daily from 1979). The CORE data used herein have been updated slightly from those described in Large and Yeager (2009). Most notably, the river runoff is now a 12-month climatology based on Dai et al. (2009).

The primary model solutions investigated are five ensemble members of the 20C fully coupled simulations of the CCSM4. They span the years 1850 through 2005 and include most of the CORE period (1948–2007). The simulations are forced with twentieth-century GHG emissions, prescribed aerosols, solar cycles, volcanic activity, and land use change (Gent et al. 2011). Consistent with increased GHGs, a late twentieth-century warming trend is present in every ensemble member. There is little ensemble spread among the differences with CORE, so

for present purposes either the ensemble mean or any one ensemble member is representative. A similar ensemble of twentieth-century simulations using CCSM3 (Collins et al. 2006) is used to document progress in reducing mean flux biases. The complete suite of fluxes that drive the CCSM4 ocean are given in Bryan et al. (1996), and details of sea ice formation and the melting of both sea ice and land ice can be found in Smith et al. (2010).

The ocean surface flux components of present interest are those for which there are CORE fluxes for comparison. They are all defined as positive into the ocean. The specific heat flux components are the net incoming solar radiation after reflection Q_S, the downward longwave radiation Q_{Ldn}, the upward longwave emission from the ocean Q_{Lup}, and the latent and sensible heat fluxes Q_E and Q_H, respectively. For some purposes it is useful to follow Large and Yeager (2012), who combine components according to the principal source of their variability. Those that vary with properties of the overlying sky, such as clouds and water vapor content combine to give an always positive sky flux:

$$Q_{\mathrm{sky}} = Q_{\mathrm{Ldn}} + Q_S. \tag{1}$$

The other components depend on PBL properties and are parameterized in terms of the differences between lower atmospheric boundary layer (ABL) and ocean surface parameters, namely, temperature ΔT, specific humidity Δq, and flow $\Delta\mathbf{U}$. We combine these into a planetary boundary layer flux Q_{pbl} that always cools the ocean. This flux and its components are computed as

$$Q_{\mathrm{pbl}} = Q_{\mathrm{Lup}} + Q_E + Q_H, \tag{2}$$

$$Q_{\mathrm{Lup}} = -\sigma\,\mathrm{SST}^4, \tag{3}$$

$$Q_E = \rho\Lambda_v C_E|\Delta\mathbf{U}|\,\Delta q, \tag{4}$$

$$Q_H = \rho c_p C_H|\Delta\mathbf{U}|\,\Delta T, \tag{5}$$

where $\rho = 1.2$ kg m^{-3} is air density, $c_p \approx 1005$ J kg^{-1} °C^{-1} is the specific heat of air, $\Lambda_v \approx 2.5 \times 10^6$ J kg^{-1} is the latent heat of vaporization, $\sigma = 5.67 \times 10^{-8}$ W m^{-2} K^{-4} is the Stefan–Boltzmann constant, and C_E and C_H are empirical bulk transfer coefficients.

The freshwater fluxes analyzed are the precipitation, P (<0), and the evaporation, E (<0 usually). The latter, and hence its mean and variability, is directly related to the latent heat flux by $E = Q_E/\Lambda_v$, so it can be regarded as a PBL water flux, and P is a sky water flux. Mean continental runoff R is considered, but not runoff variability, because only the variability of gauged rivers is directly included in the CORE data. Following CORE, runoff is distributed over ocean coastal regions adjacent to river mouths, so the runoff is expressed as an ocean surface flux over these regions. Combining PBL and sky fluxes gives the air–sea fluxes

$$Q_{\mathrm{as}} = Q_{\mathrm{sky}} + Q_{\mathrm{pbl}}, \tag{6}$$

$$F_{\mathrm{as}} = P + E, \tag{7}$$

where each is the sum of two numbers of similar magnitude, but opposite sign. In the presence of sea ice, only the fluxes through the open water fraction contribute to these air–sea fluxes. Exchanges with sea ice are not considered here because of the lack of comparable observations.

3. Mean state

In this section ''present-day'' mean air–sea fluxes Q_{as} and F_{as} and runoff R are computed as time averages over the last 20 years of both the CCSM3 (1980–99) and CCSM4 (1986–2005) 20C ensemble means. The combined temporal plus ensemble averaging is preferred to reduce natural variability in the model, but results from any single member would be very similar. The main biases discussed here exceed the ensemble spread and so are not likely to be attributable to model internal variability. For comparison, present-day CORE air–sea fluxes are averaged over the 20-yr time period corresponding to the model average, but with only one observed realization, there is no reduction of natural variability by ensemble averaging in CORE.

Differences between CCSM3 and CCSM4, particularly in the present-day fluxes and SST, are expected because of a number of factors. There is different physics, although improved or extended physics does not necessarily improve simulations because of compensating errors. The resolution and dynamical core of the coupled atmosphere model and vertical resolution of the ocean model are different. The sea ice cover is not the same (Jahn et al. 2012; Holland et al. 2006). The compared time periods are not identical since CCSM3 simulations ended in 1999 and CCSM4 in 2005. The tuning and spinup procedures were distinctly different (Gent et al. 2011). CCSM3 is tuned to match 1990 GHG and aerosol conditions, when observations are best, but the model's top of the atmosphere radiation balance is unrealistic. In contrast, CCSM4 is tuned to match more poorly known 1850 reconstructions, but has a more defensible top of the atmosphere radiative balance.

Comparison of the present-day means of heat and freshwater flux fields from the CCSM4 to CCSM3 (Large and Danabasoglu 2006; Hack et al. 2006; Collins et al.

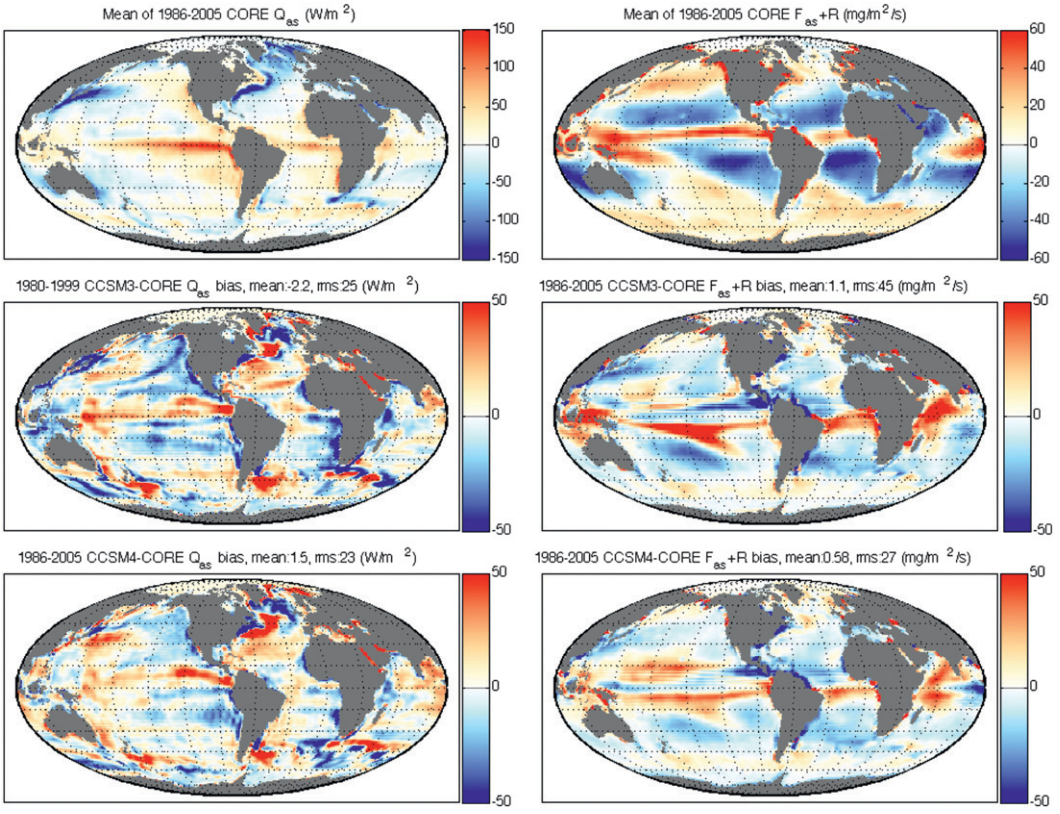

FIG. 1. (top) CORE (left) total air–sea heat flux and (right) total freshwater flux (air–sea + runoff) into the ocean. Also shown are biases in the present-day mean of these fluxes from the (middle) CCSM3 and (bottom) CCSM4 20C ensemble means. Units: W m^{-2}. The increment in latitude is 15°.

2006) shows improvement, although many known biases are still present and are sometimes worse (Fig. 1). The global present-day mean bias for the Q_{as} flux indicates an overall increase of heat flux into the ocean with the transition from CCSM3 to CCSM4 with a global mean bias value of -2.2 W m^{-2} in CCSM3 and 1.5 W m^{-2} in CCSM4. The root-mean-square error (RMSE) has also decreased slightly from 25 W m^{-2} in CCSM3 to 23 W m^{-2} in CCSM4. The net $F_{as} + R$ bias has also improved from a global mean bias of 1.1 mg m^{-2} s^{-1} in CCSM3 to 0.58 mg m^{-2} s^{-1} in CCSM4. The RMSE for freshwater flux has greatly decreased from 45 to 27 mg m^{-2} s^{-1}. The most notable improvements in the present-day mean $F_{as} + R$ are a reduction of positive biases in the tropical South Pacific, tropical Atlantic, Maritime Continent, and western Indian Ocean (Fig. 1, right panels). Improvements in mean heat flux include a reduction of biases in the north tropical Atlantic basin, central to western equatorial Pacific, and western and equatorial Indian Ocean (Fig. 1, left panels).

The zonal mean of biases and RMSE for Q_{as} and $F_{as} + R$ are shown in Fig. 3. Improvements in RMSE are quite large in $F_{as} + R$ from approximately 10° to 30°S and slight improvement for most of the Northern Hemisphere south of 60°N (solid lines). Differences in mean biases of $F_{as} + R$ (dashed lines) are not correlated to the RMSE and do not span large latitudinal ranges. A similar widespread reduction of RMSE is noted in Q_{as} from approximately 30°S to 30°N (solid lines). An overall increase of Q_{as} mean bias occurs CCSM4 from CCSM3 in the latitudinal range of 30°S to 40°N (dashed lines), thus improving negative biases and causing positive biases to be worse.

The majority of the freshwater flux improvement results from large improvements in precipitation biases, which are reflected in surface salinity biases (see Danabasoglu et al. 2011). Improvements in the atmosphere model convection scheme (Richter and Rasch 2008) lead to improvements in the statistics of precipitation extreme events; however, some mean biases remain. The erroneous double intertropical convergence zone (ITCZ) south of the equator still exists and there is an exacerbation of the positive precipitation bias associated with the ITCZ north of the equator in the Pacific Ocean (visible in the $F_{as} + R$ plot of Fig. 1). In general, the central to western equatorial and midlatitude

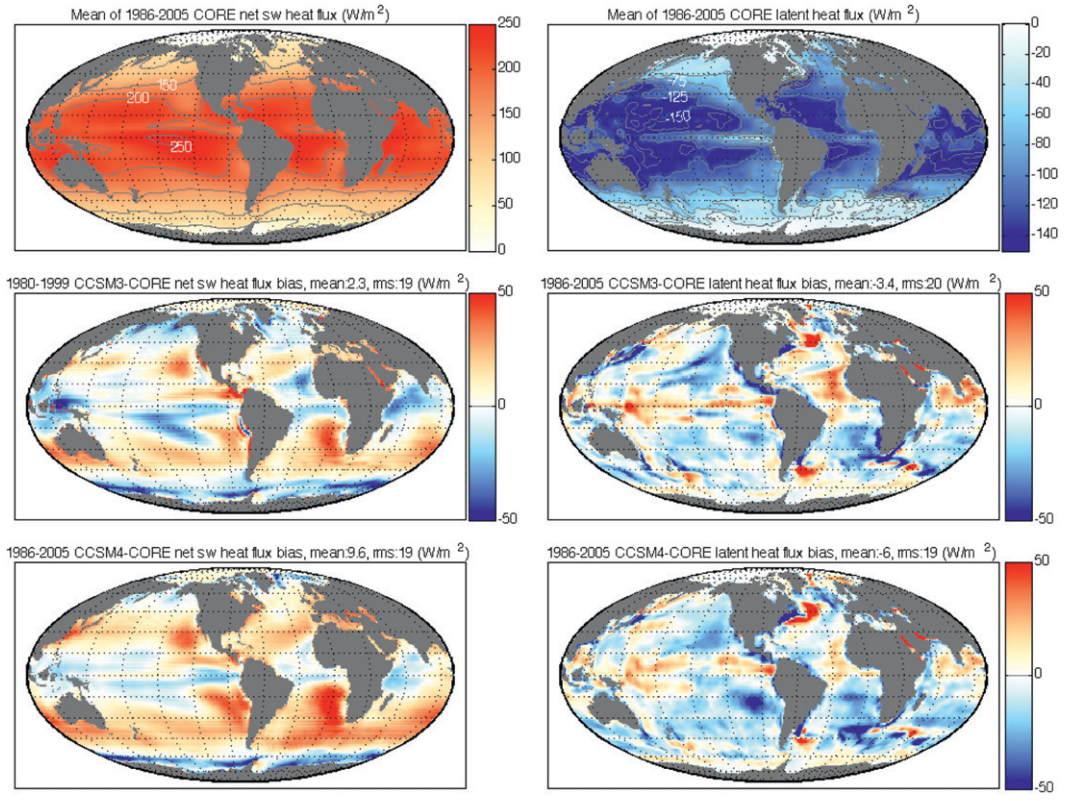

FIG. 2. As in Fig. 1, but for (left) net shortwave heat flux and (right) latent heat flux.

Pacific go from too saline (CCSM3) to too fresh (CCSM4) (Danabasoglu et al. 2011). Present-day mean precipitation biases (not shown) also indicate reduced precipitation in CCSM4 from CCSM3 in Indonesia and a better representation of precipitation in the South Pacific convergence zone (SPCZ). In CCSM3, the SPCZ extends to 130°W whereas in the CCSM4 it extends only to 160°W.

Examination of the individual components of the air–sea heat flux reveals that the majority of improvement in Q_{as} in the tropics is due to a reduction of biases in latent heat flux (Figs. 2 and 3). Although the global mean bias increases from -3.4 W m^{-2} in CCSM3 to -6 W m^{-2} in CCSM4, there are significant improvements in regions of largest error. The largest improvements in Q_{as} are in the tropics of all ocean basins. These are the same regions with largest improvement in Q_E, most notably the tropical North Atlantic and Maritime Continent region. All of these regions also have improved SST biases as well (Danabasoglu et al. 2011) reflecting the connection between evaporation and SST. With the exception of the equatorial region, the zonal mean of bias and RMS errors do not reflect these improvements (Fig. 3).

Net shortwave radiation is degraded in the transition from CCSM3 to CCSM4 (Figs. 2 and 3) with a global mean bias increase from 2.3 to 9.6 W m^{-2}. With a nearly uniform increase, the result is that negative biases in CCSM3 are reduced and positive ones made even worse in CCSM4. Although the zonal mean of biases reflects the degradation at almost all latitudes, the zonal average of RMS shows improvement in CCSM4. This reflects the considerable compensation of regional error that can be hidden when regionally averaging.

The CCSM4 present-day (1986–2005) regionally averaged flux components are compared to a collection of flux datasets [compiled by Roske (2006)] in Figs. 4 and 5. All data are presented here as differences from the regional mean of the CORE flux components. The ensemble mean difference is displayed as an asterisk, and the range in differences of various observational datasets is indicated by the vertical line. Note that this line does not represent error, but rather the range of differences of the Roske (2006) datasets from CORE. The purpose of this exercise is to illustrate how the CCSM4 data compare to a variety of flux datasets and how large their range is.

For many regional components, the observational range is quite large, thus making it difficult to unambiguously test model performance; however, the main conclusion to be drawn from these plots is that the model fluxes lie

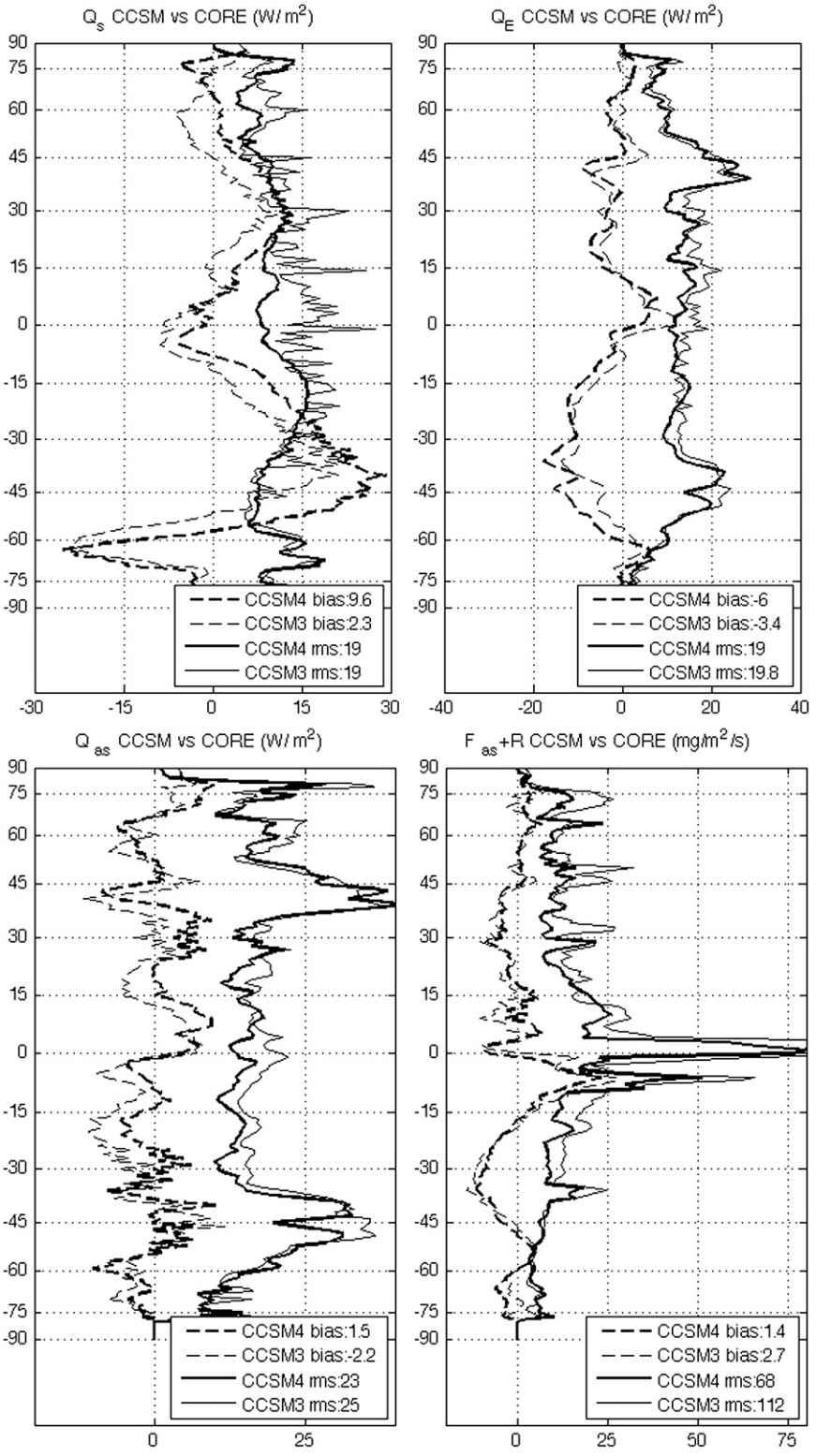

FIG. 3. Zonal average of the present-day mean bias and root-mean-square error (RMSE) values calculated from (top) Fig. 2 and (bottom) Fig. 1. The bold line is the CCSM4 20C ensemble mean and the thin line is from CCSM3. Dashed lines are the zonal mean bias and solid lines are the zonal mean RMSE. Units: W m^{-2}.

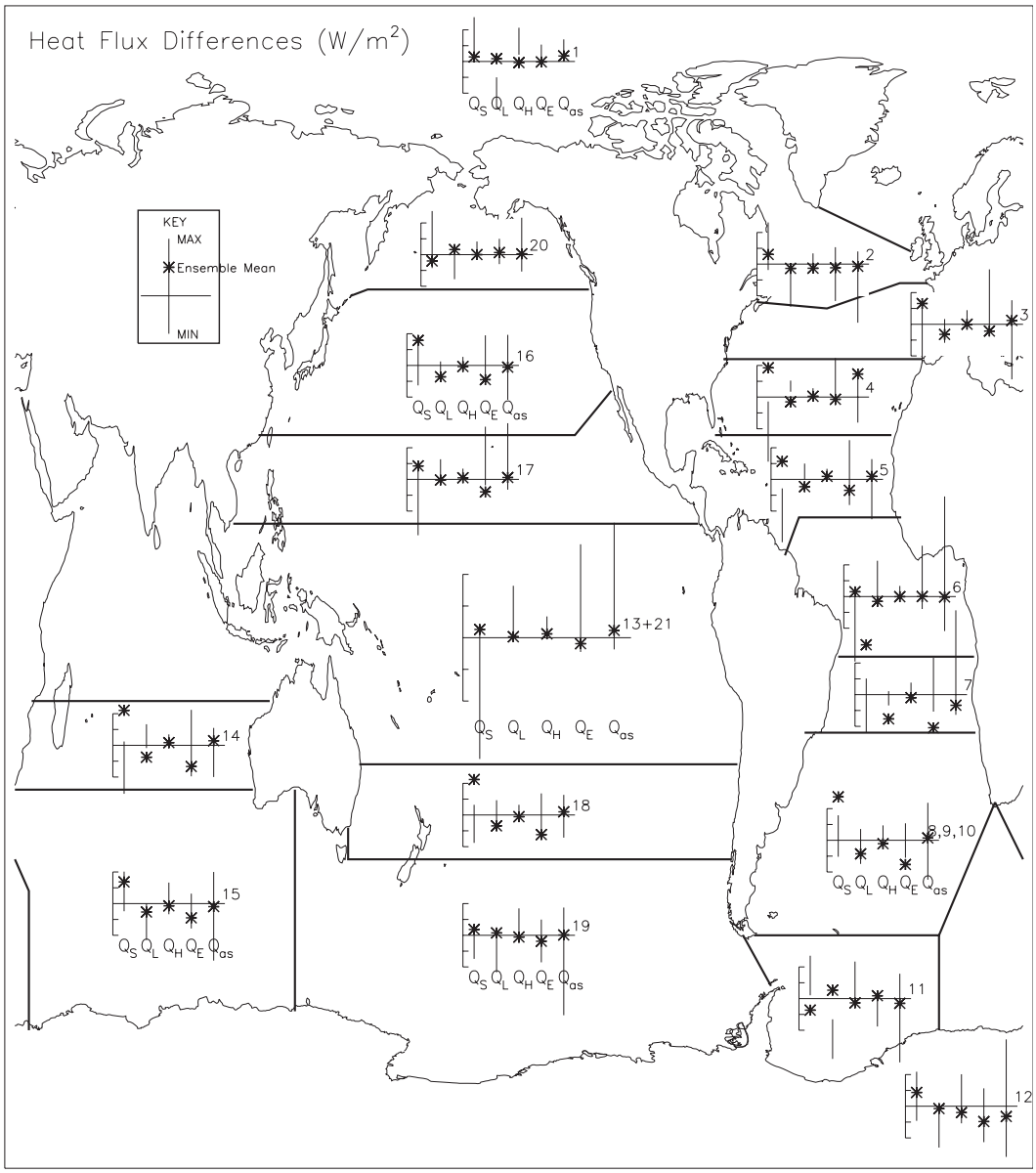

FIG. 4. Regional averages of the CCSM4 20C ensemble mean heat flux components differenced with the CORE fluxes (asterisk). Region 13 + 21 includes both the equatorial Pacific (13) and Indian Ocean (21) basins. The vertical line represents the range of observed datasets compiled by Roske (2006) differenced with CORE. The region numbers are given at the right side of each x axis. The components are net shortwave heat flux (Q_S), net longwave heat flux (Q_L), sensible heat flux (Q_H), latent heat flux (Q_E), and total air–sea heat flux (Q_{as}) following the labeling convention in Large and Yeager (2009). The units are W m^{-2} and the y-axis increment is 10 W m^{-2}.

within the range of observed values for most flux components and regions. In many cases, the errors in individual components, when averaged together, create a Q_{as} flux that is very close to the CORE Q_{as} flux. Examples are regions 2, 6, 8–10, 16, 17, 19, and 20 corresponding to most of the Atlantic basin and the North and South Pacific in Fig. 4. Similarly, biases are found in most regions in each component of net longwave radiation (up and down; not shown) compared to CORE; however, the net longwave radiation is similar to the net CORE longwave, indicating that biases in one component are compensated by the other. Additionally, heat flux biases tend to be localized and not as zonal as the freshwater biases so that, when averaged over one of the regions, the mean bias is small.

From Figs. 4 and 5, we can see which components may be responsible for the mean biases shown in Fig. 1. The regions in which the shortwave heat flux (Q_S) is outside

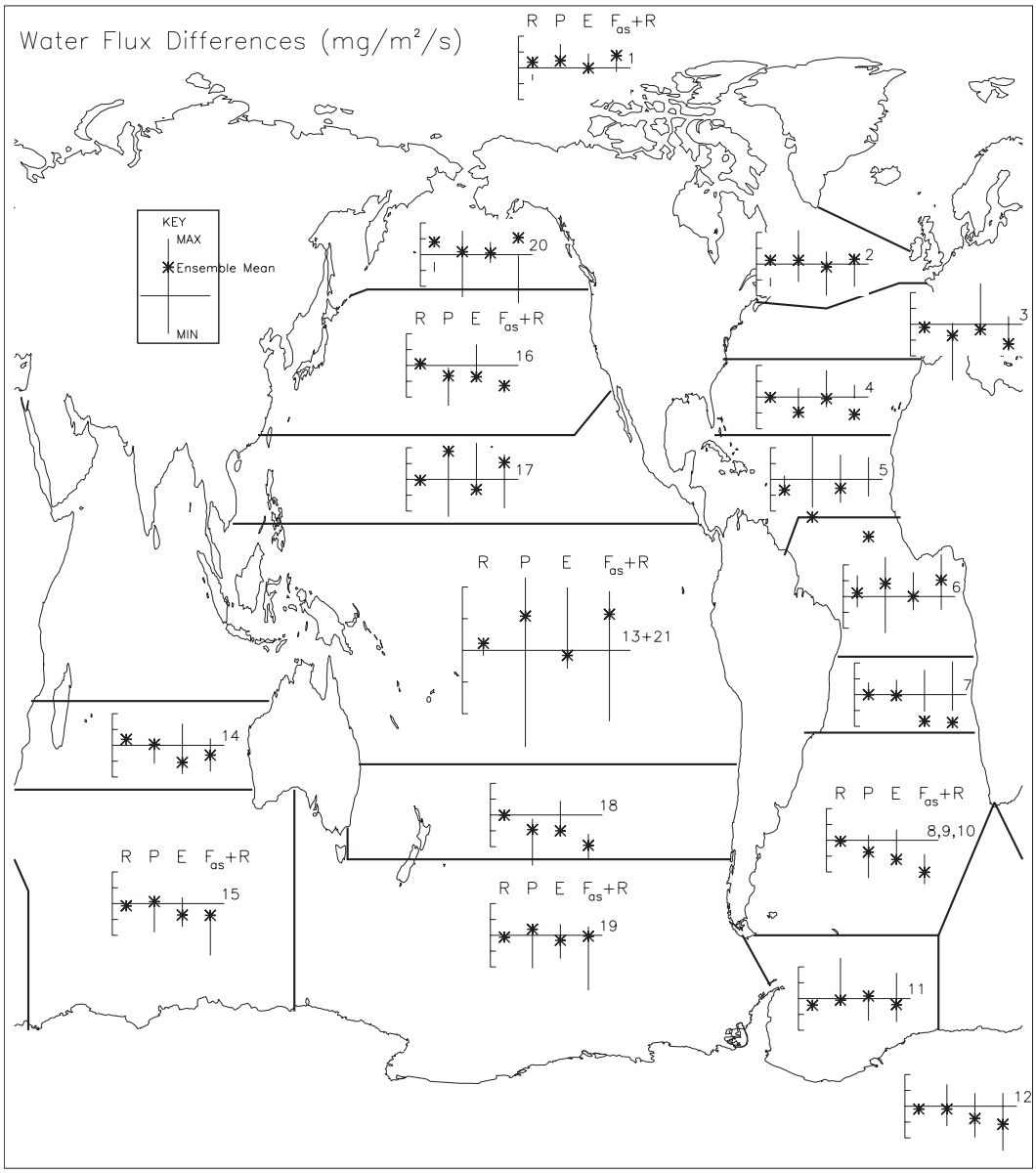

FIG. 5. As in Fig. 4, but for freshwater flux. The components are runoff R, precipitation P, evaporation E, and total air–sea freshwater flux F_{as} plus runoff. The units are mg m^{-2} s^{-1} and the y-axis increment is 5 mg m^{-2} s^{-1}.

the range of known observations (regions 4, 5, 7–11, 14, and 18) correspond to mid- to high-latitude portions of each ocean basin where the largest errors are found in the mean (Figs. 2 and 3). The large positive bias compared to CORE in precipitation due to the double ITCZ and enhanced precipitation over Indonesia is evident in the equatorial Indian and Pacific basins (region 13 + 21) of Fig. 5 and contributes to the large positive biases in the mean freshwater flux. Additionally, the North Atlantic (regions 3–5) exhibits a negative bias in precipitation over the North Atlantic basin, which leads to an overall negative mean freshwater bias seen in Fig. 1. Additionally, the runoff errors of the Congo and Baffin Bay/Canadian Archipelago, due to an excess of precipitation over the continents (Danabasoglu et al. 2011), contribute to the positive bias in regions 6 and 1, respectively. Improvement in Nile River runoff has reduced the freshwater flux into the Mediterranean Sea, improving the local freshwater bias. Although not reflected in the runoff bias of region 6, runoff from the Amazon River is too small.

The air–sea heat flux balances in the 21 regions in Fig. 4 can be classified in two ways: regions in which all heat flux components are in good agreement with the CORE data and within the range of observation-based estimates and

regions in which there is a significant excess of solar radiation. The classification does not necessarily point out regional behavior, but is presented as a way to simplify our summary of results. The first classification includes the Arctic, subpolar and equatorial Atlantic, equatorial Indian and Pacific, north tropical and northern North Pacific, and Southern Ocean regions at the southern end of the Atlantic and Pacific basins (regions 1, 2, 6, 12, 13 + 21, 17, 19, and 20). These represent approximately 57% of the total ocean surface. Included in second classification are the Atlantic basin, excluding the subpolar and equatorial regions, the Indian Ocean, and the North and South Pacific (regions 3–5, 7, 8–11, 14–16, and 18). These represent approximately 40% of the ocean surface. The excess solar radiation is partially compensated by a deficit in the net longwave radiation. Such a compensation is expected when cloud cover is the problem, as too little cloud allows too much solar energy to reach the surface, but emits less longwave radiation down to the surface. Only in the North Atlantic (region 4) is the compensation weak enough and the other fluxes near enough to the CORE data that the solar bias is reflected strongly in Q_{as}. In the other regions, the longwave compensation plus a deficit in latent heat flux puts Q_{as} in much better agreement with CORE. With the exception of the North Atlantic, tropical South Atlantic, and a portion of the Southern Ocean south of the tip of Africa (regions 4, 7, and 12), the agreement is good, so, in combination with the first classification, the agreement with regional CORE Q_{as} extends over 92% of the ocean, and thus the implied heat fluxes are similar and, as with CORE, in good agreement with heat transports inferred from ocean observations.

The freshwater ($F_{as} + R$) flux balances in the 21 regions in Fig. 5 can be classified in three ways based on precipitation. In general, there is no compensation between components because precipitation and runoff depend largely on remote evaporation and precipitation, respectively, while evaporation is directly related to the latent heat flux. The first grouping includes the Arctic, subpolar North Atlantic, Indian, northern North Pacific, and Southern Ocean south of the Atlantic and Pacific basins (regions 1–2, 11, 14, 15, 19, and 20). These account for 29% of the ocean surface. These regions have good agreement with CORE, most notably in precipitation, and are within the range of observations in ice-free regions, with the exception of the Arctic, subpolar North Atlantic, and northern North Pacific (regions 1, 2, and 20) where runoff is out of range. Excess runoff into the northern North Pacific (region 20) makes the $F_{as} + R$ water flux agree less well with CORE than evaporation and precipitation would suggest. The second grouping includes the equatorial Atlantic, Pacific, and Indian basins and the north tropical Pacific (regions 6, 13 + 21, and 17), which are characterized by significantly greater precipitation than CORE and hence greater $F_{as} + R$ flux. However, these regions include the ITCZ areas and could be considered in the first classification because the high mean rainfall rates mean that the percentage differences are small and the range of estimates over these regions is very large. Finally, the third grouping includes the North Atlantic, South Atlantic, North Pacific, tropical South Pacific, and Southern Ocean south of the Atlantic basin (regions 3, 4, 8–10, 12, 16, and 18) in which precipitation tends to be lower than CORE and the sum of the other flux components reinforces this bias, such that the $F_{as} + R$ flux is significantly less than CORE. The south tropical Atlantic (region 7) is atypical because the only significant bias is evaporation. At the other extreme is the north tropical Atlantic (region 5), where precipitation is very low, runoff and evaporation are also low, and the $F_{as} + R$ flux is more than 30 mg m^{-2} s^{-1} (75 W m^{-2}) less than CORE.

4. Annual and interannual variability

To begin routine assessment of the model's ability to represent the annual and interannual variability of the air–sea fluxes, the variance of each flux component within annual and interannual frequency bands (from monthly mean data) is compared to the variance contained in each corresponding CORE flux component. As this is the initial assessment of these errors, the discussion focuses on CCSM4. Also, the discussion herein is more a description of the present errors than a prescription for their correction, as the underlying mechanisms and sensitivities to model configuration are presently poorly known.

Comparisons of annual and interannual variability in this section include the entire suite of five ensemble members for the period 1984–2005.[1] The time period is based on when radiation data are available from CORE; however, using the entire available time period of 1949–2005 did not alter the results significantly. The data are bandpass filtered to retain either 9–15-month (annual) or 2–7-yr (interannual) time scales.[2]

Figures 6 and 7 map the standard deviation of CORE and the ratio of the variance in the CCSM4 versus CORE

[1] The ensemble-mean variance is calculated by first taking the temporal variance of each ensemble member and then taking the mean of the temporal variances across the ensembles. The ensemble spread of variance is the spread across the ensemble of the temporal variances from each ensemble member.

[2] Bandpass filtering is done with a 10th-order Butterworth filter. Limits given are the cutoff frequencies for the −3-dB point (variance per frequency reduced by half) below the passband value.

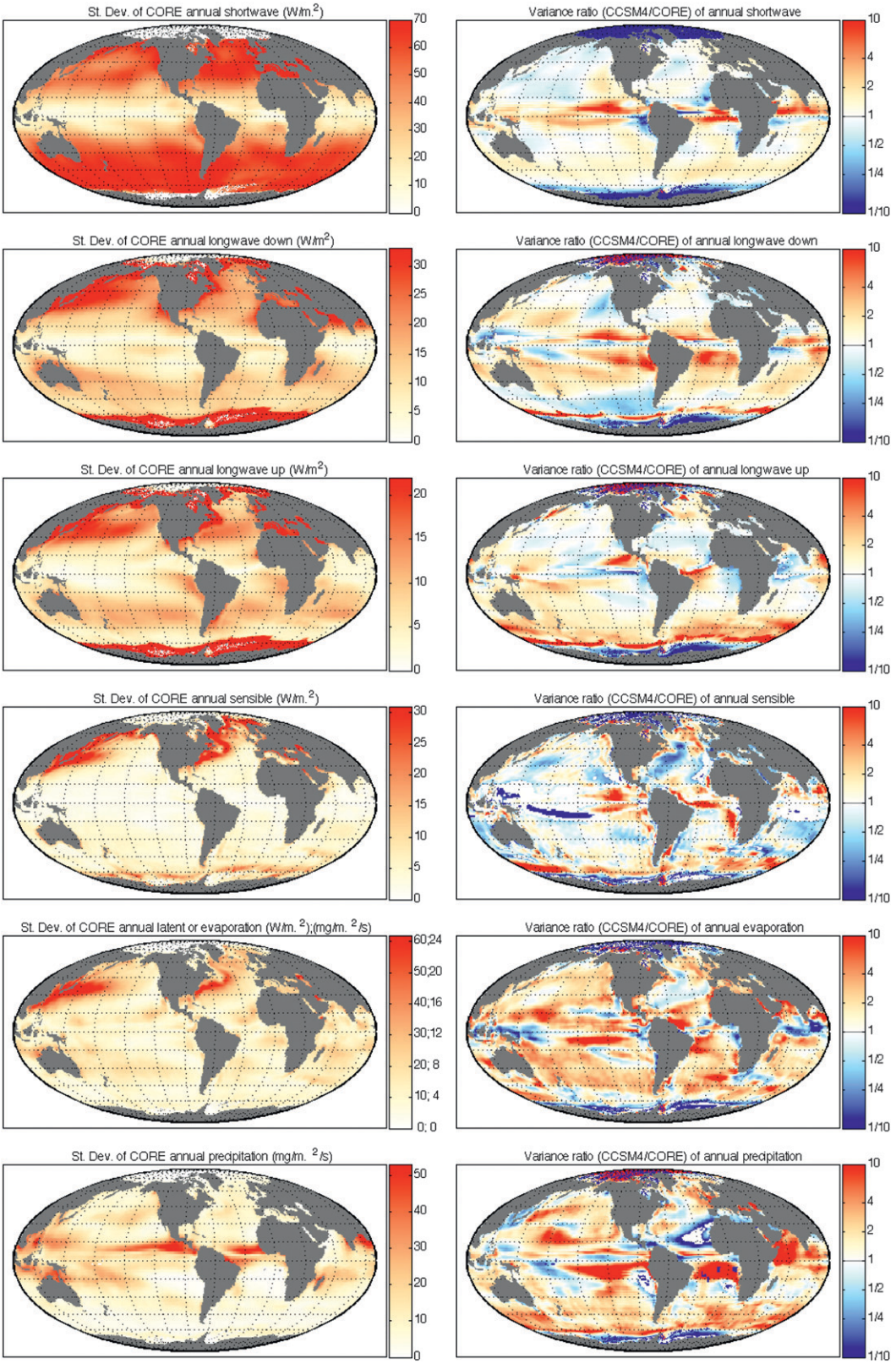

FIG. 6. Maps of (left) CORE standard deviation and (right) the ratio of CCSM4 to CORE variance for the 9- to 15-month bandpassed variability for 1984–2006. Only regions where CORE standard deviation falls outside of the ensemble range of standard deviations are plotted in the ratio figures. The increment in latitude is 15°.

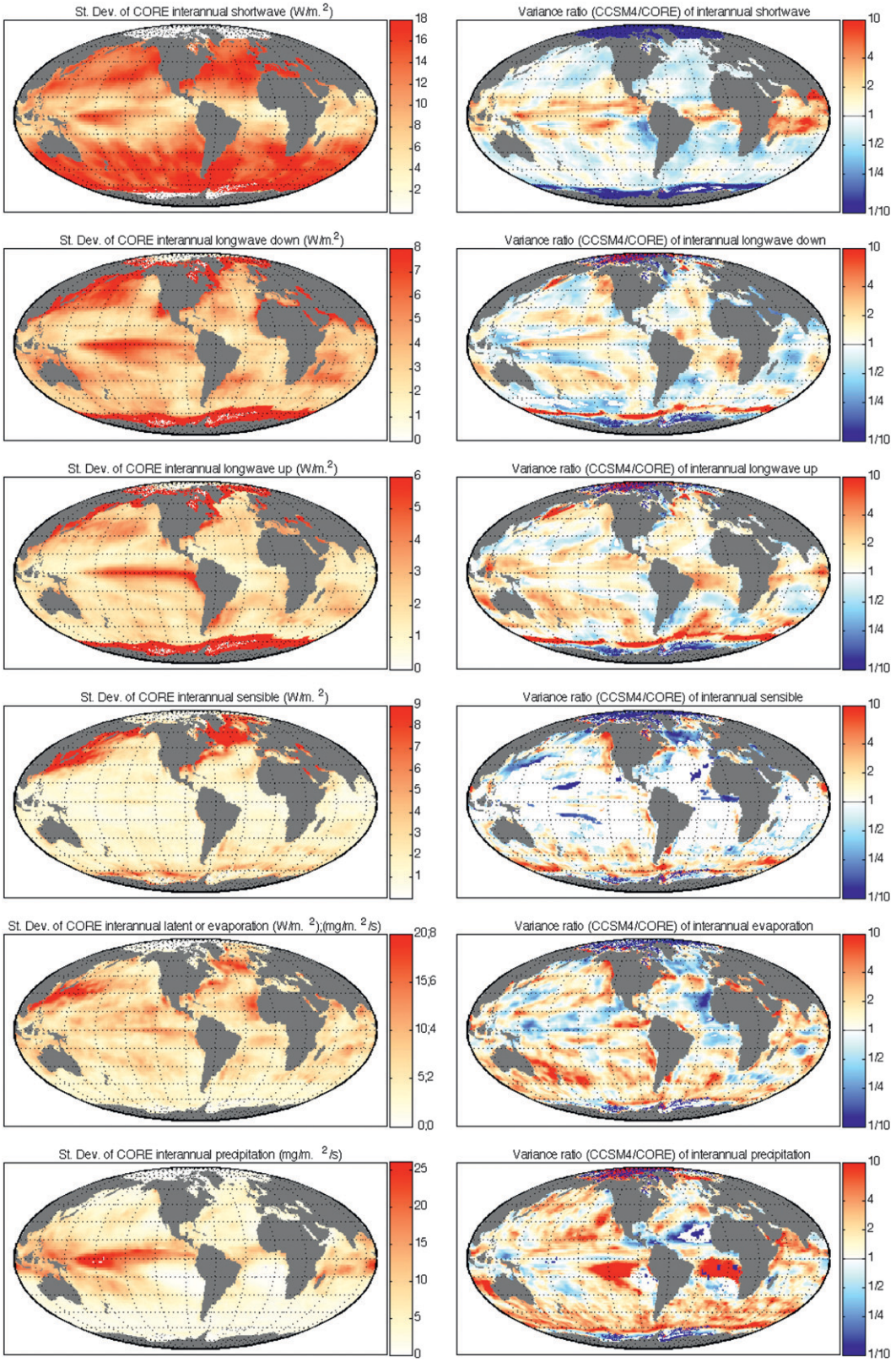

FIG. 7. As in Fig. 6, but for the 2–7-yr variability.

on annual and interannual time scales. Since the desired variance ratio is 1, the color bar is chosen to show where CCSM4 has a deficit of variability (blue) or an excess (red). These nonunity variance ratios will loosely be referred to as ''errors'' in variance. Because of the likelihood of sea ice errors dominating flux errors at polar latitudes, the contours and coloring in these figures focuses on the errors from 40°S to 40°N. Many of the errors in annual and interannual variability are easily recognizable as related to known biases in the mean state such as western boundary current and double ITCZ errors (Figs. 1 and 2). However, some of the errors in the annual and interannual bands differ in location and magnitude relative to each other and the mean errors.

The patterns of variability are similar in the data and model (model total variance is not shown) in that in the regions where variance in CORE data is larger, the variance in CCSM4 is generally also larger (along with a larger ensemble spread, typically). However, it is rarely the case that the ensemble variance spread includes the CORE variance regionally. Only those values exceeding ensemble variability limits are shaded in Figs. 6 and 7, and the majority of the globe is shaded. Furthermore, the variance ratios of CCSM4 to CORE often reveal regions where variances are a factor of 10 in error. Thus, in many regions and for most fluxes, the CCSM4 is systematically biased such that the limits of ensemble spread cannot explain annual and interannual time scales.

The most poorly modeled heat flux, in terms of the variability error magnitude, is the shortwave heat flux (Figs. 6 and 7, top panels). Shortwave variability involves many factors, including the effects of clouds, absorption, albedo, and top of the atmosphere radiation; errors in modeling these processes lead to $O(30\ \mathrm{W\ m^{-2}})$ variability errors. The annual shortwave error is focused on the eastern sides of each basin and is probably related to the modeled monsoons. The interannual shortwave error is spread across the tropics, indicating that lower-frequency variability of the ITCZ is more zonally uniform.

Other heat flux errors are generally smaller in magnitude, although they may be regionally important. Upward and downward longwave errors are quite different over the annual and interannual bands. Downward longwave radiation variability errors share similar patterns to the shortwave errors, consistent with erroneous cloud variability. Many studies indicate that sea surface temperature variability can be improved in future models, with resolution (McClean et al. 2012) or parameterizations (Fox-Kemper et al. 2011). If achieved, then upward longwave radiation would also be improved. Currently, upward longwave variability is overly strong for most of the globe on interannual time scales and for most of the Southern Hemisphere outside of the equatorial region and in high latitudes on annual time scales, which suggests that sea surface temperature variability is excessive in these regions as well. Sensible heat flux variability errors are concentrated above regions of strong oceanic temperature gradients, but annual and interannual errors differ substantially. Sea surface temperature variability is regionally important as it strongly influences atmospheric patterns (Kaspi and Schneider 2011).

The patterns of annual and interannual variability in latent heat flux and precipitation in the CORE data are familiar (e.g., Xie and Arkin 1997) in that they resemble familiar features such as El Niño–Southern Oscillation, ITCZ migration, and western boundary current variability, but the patterns of *erroneous* variability are not. Many of the evaporation and precipitation variability errors appear in the mean differences (Figs. 1 and 2) albeit with different magnitudes. Intriguingly, some errors appear in the mean difference but not in the annual or interannual variability errors, such as the positive equatorial errors in latent heat flux in all basins in CCSM4. Other errors, such as the latent heat flux error in the western boundary currents, appear in all frequency bands. Latent heat flux variability is mainly in the western subtropics of each basin in regions that contain storm tracks. The precipitation variability is dominated at all time scales by the ITCZ, with its associated rainy seasons. CCSM4 variability is biased such that latent heat flux and hence precipitation standard deviations are too large in virtually all regions and well beyond the CCSM4 ensemble spread. In the annual band, the dominant latent heat flux errors are in the East African, southwest Indian, and American monsoon regions and western boundary current regions, both of which are better simulated in the interannual band. In precipitation, the erroneous double ITCZ dominates both annual and interannual variability errors, although there is a sizeable storm track signal in the annual band precipitation as well.

In Figs. 6 and 7, the only variance errors colored are those where the CORE variance lies outside of the CCSM4 ensemble spread. However, this coloring provides only a rough estimate of error significance. Furthermore, the CORE dataset involves interpolation and is thus likely to be inaccurate on a gridpoint by gridpoint level. For a quantitative estimate of significance of variance errors, we employ another statistical measure, the wavelet probability analysis (WPA) hypothesis testing procedure (Stevenson et al. 2010). WPA was designed for use with climate index data, where the variability is generally not normally distributed (e.g., El Niño indices), nor are the variances distributed according to chi-squared statistics (although WPA will reproduce these distributions if they do occur). The WPA hypothesis testing procedure is applied to a set of climate indices

TABLE 1. Results from the wavelet probability analysis (WPA) operating on data in the 9–15-month band. Values given are the $(1 - p)$ level; the p value is the minimum confidence level that variability of the coupled model and CORE in a particular region are different (the null hypothesis is that they agree). Therefore, the lower the value is, the less sure we are of model variability error (i.e., $1 - p = 0.9$ corresponds to a 90% likelihood of disagreement). Time periods used for this analysis are as follows: P and Q_H (1979–2005), Q_E (1949–2005), and Q_S and Q_L (1984–2005). Values ≥0.90 are in bold.

Region	Q_E	P	Q_H	Q_L	Q_S
1	0.85	**1.00**	**0.97**	0.71	**0.99**
2	**1.00**	0.73	**0.94**	0.01	0.06
3	0.78	0.00	0.76	0.32	0.87
4	0.11	0.35	**1.00**	**0.98**	0.01
5	**1.00**	**0.92**	0.04	**1.00**	0.21
6	**0.93**	**1.00**	**1.00**	**0.99**	**1.00**
7	**1.00**	**1.00**	0.79	**1.00**	**0.97**
8	0.86	0.68	0.26	0.80	**1.00**
9	**1.00**	0.80	0.25	**1.00**	**1.00**
10	0.84	**1.00**	0.19	**1.00**	**1.00**
11	**0.99**	0.19	0.05	**1.00**	**1.00**
12	0.23	0.17	0.83	**1.00**	0.09
13	**1.00**	0.53	**1.00**	0.75	0.53
14	**0.99**	0.22	**0.92**	**1.00**	**0.97**
15	0.75	**0.99**	0.42	**1.00**	**1.00**
16	**1.00**	**0.99**	0.16	0.23	0.13
17	**1.00**	0.89	0.55	**1.00**	0.19
18	**1.00**	**1.00**	0.82	**1.00**	**1.00**
19	0.33	0.74	0.32	0.23	0.51
20	**0.90**	0.15	0.42	0.05	**1.00**
21	**0.91**	0.78	0.50	**1.00**	0.86

here: the annual and interannual scale-averaged wavelet power (Torrence and Compo 1998) of the regionally averaged fluxes. The averaging regions chosen are the same as those indicated in Figs. 4 and 5. The WPA hypothesis test estimates the probability that the variability of two time series differs. In this particular case, WPA is used to find the likelihood that the joint probability distributions of subintervals of the CCSM4 simulations differ from those created from CCSM4–CORE comparisons, which is an indication that the regionally averaged, bandpassed variability has distinct statistical properties in CCSM4 versus CORE. The time period for the flux components used for the WPA varies depending on the time period of available data for CORE.

Tables 1 and 2 give the WPA significance values for each flux component for annual and interannual variability, respectively. Each entry corresponds to the significance of CCSM4–CORE differences and are reported as $(1 - p)$, where p is the probability. For example, a value of 0.9 indicates that there is a 90% likelihood that the variability of the bandpass-filtered, regionally averaged fluxes in CORE and the CCSM4 ensemble differ. Likewise, a low value indicates that the CCSM4 and CORE do not differ significantly. Values greater than 0.90 are in bold in the tables to highlight the disagreement.

TABLE 2. As in Table 1, but for data in the 2–7-yr band.

Region	Q_E	P	Q_H	Q_L	Q_S
1	0.11	0.19	0.26	0.18	0.10
2	0.17	0.07	0.03	0.07	0.27
3	0.13	0.27	0.01	0.02	0.39
4	0.30	0.29	0.08	0.10	0.25
5	0.26	0.16	0.65	0.22	0.41
6	0.17	0.80	0.21	0.01	0.24
7	0.21	0.31	0.35	0.33	0.06
8	0.13	0.44	0.28	0.02	0.34
9	0.21	0.28	0.15	0.48	0.30
10	0.30	0.31	0.04	0.36	0.28
11	0.09	0.13	0.04	0.26	0.07
12	0.22	0.29	0.22	0.01	0.07
13	0.13	0.09	0.58	0.45	0.05
14	0.03	0.29	0.09	0.28	0.20
15	0.29	0.03	0.27	0.08	0.19
16	0.01	0.07	0.29	0.16	0.11
17	0.17	0.31	0.08	0.12	0.19
18	0.27	0.36	0.33	0.42	0.20
19	0.19	0.41	0.02	0.25	0.22
20	0.16	0.31	0.07	0.22	0.17
21	0.26	0.32	0.34	0.25	0.10

The most striking result is that model performance is much worse for annual variability than for interannual variability. Virtually all of the regions show strongly significant errors in the annual band in all flux components with almost half of the values in the table above the 90% confidence level for CCSM4–CORE disagreement. The tropical Atlantic (regions 6 and 7) is noteworthy for poor simulation of annual variability in flux components, and the components with erroneous variability in the most regions are Q_E, Q_L, and Q_S.

For interannual variability, all of the WPA values in Table 2 are below the 90% confidence level with only one value above 80%. This agreement may be partially due to the short observational record available for some of the CORE flux components, but it still suggests that the focus for model flux variability improvement should be the seasonal cycle.

5. SST and heat flux trends and feedbacks

Large and Yeager (2012) analyzed the CORE interannual fluxes from 1984 through 2006 and found a global SST increase of 0.28°C with a corresponding change in Q_{as} of -9.1 W m^{-2}. The conundrum they faced was how SST could continue to rise while the surface heat flux into the ocean was decreasing. Their indirect inference was that as the SST warmed there was less cooling due to mixing with colder subsurface water, perhaps because of increased stability of the upper ocean.

Here we seek to understand how the modeled CCSM4 heat fluxes behave during a time of upward trending SST and if this behavior is different from the CORE data. These questions are addressed in one particular CCSM4 ensemble member that was chosen because its 25-yr (1981–2005) global SST change of Δ_{25}SST = 0.26°C matches that of the Hurrell et al. (2008) dataset used to compute CORE fluxes. This ensemble member also has the best representation of the observed late twentieth-century decline in Arctic sea ice (Jahn et al. 2012). We use Δ_{25} to indicate the change over the 25-yr period 1981–2005 determined from a linear regression of the 25 annual means (see Fig. 8, top left) at every ice-free spatial point. The range of Δ_{25}SST from the other four ensemble members is 0.36° to 0.39°C and they all display qualitatively similar flux trend behavior.

Sea surface temperature has been observed to be warming since the 1970s in all ocean basins (e.g., Hurrell et al. 2008). The negative feedback on the air–sea flux from such a warming is given by (Large and Yeager 2012)

$$\frac{\partial Q_{\rm pbl}}{\partial \rm SST} = -4\sigma \rm SST^3 - |\Delta \mathbf{U}|\rho\Lambda_v C_E \frac{\partial \rm SS}{\partial \rm SST} - |\Delta \mathbf{U}|\rho c_p C_H, \quad (8)$$

where SSq is the saturation humidity at the sea surface and the three terms on the right-hand side (rhs) arise from $Q_{\rm Lup}$ (3), Q_E (4), and Q_H (5), respectively. Since each of the rhs terms is negative, the effect of a warming SST on $Q_{\rm pbl}$ is to impede further warming by the boundary layer fluxes (Large and Yeager 2012).

We quantify flux trends by the 25-yr changes (Δ_{25}) corresponding to the Δ_{25}SST warming trend, and thus the word ''feedback'' indicates the trend of the fluxes accompanying the trend in SST. In other words, does the trend of the flux component enhance or diminish the upward SST trend? A quantitative measure of the overall feedback is the PBL coupling coefficient

$$C_{\rm pbl} = \frac{\Delta_{25} Q_{\rm pbl}}{\Delta_{25} SST}, \quad (9)$$

yielding the change in PBL heat flux per unit change in SST over the 25 years, therefore given in W m^{-2} °C^{-1}. All CCSM4 or CORE ocean cells with a nonzero ice concentration are excluded.

a. The twentieth century in CCSM4

Figure 8 presents the time series of annual mean SST and heat flux components averaged over the ice-free ocean. From 1850 to the 1920s SST and flux components are relatively steady, apart from the transient responses to volcanic eruptions in 1883 (Krakatua) and 1902 (Santa Maria). Otherwise, $Q_{\rm as}$ vacillates around 2.8 W m^{-2}. The years from the 1930s onward are characterized by increasing SST and trends in the heat flux components. At least some of the SST increase is presumably a response to greater concentrations of GHGs, including water vapor, in the model atmosphere, leading to an increase in $Q_{\rm Ldn}$ by about 6 W m^{-2}. This increase is only partially balanced by less shortwave radiation. As expected from the first term on the rhs of Eq. (8), the 0.5°C increase in SST leads to a change in $Q_{\rm Lup}$ of about −3 W m^{-2}. According to Eq. (8), both Q_H and Q_E would decrease due to an increase in SST, but what we see in the CCSM4 is only a small negative trend in Q_E and a positive trend in Q_H. As a result, the overall air–sea heat flux into the ocean $Q_{\rm as}$ increases so that there is less negative feedback to impede the warming of SST in the CCSM4 ocean.

b. CCSM4 versus CORE; 1981–2005

There are differences in the mean values of the SST and all the heat flux components between the CCSM4 and CORE (Fig. 8). In this section however, we are concerned with how the trends in the heat flux feed back onto the SST trend in the late twentieth century. Observed SST is increasing over the late twentieth century and CCSM4 shows similar behavior; however, the trends of some heat flux components have very different behavior. The largest differences between CORE and CCSM4 are in $Q_{\rm Ldn}$ and Q_H, which exhibit opposite sign trends between CORE and CCSM4. Furthermore, although acting as a cooling term in both datasets, the Q_E trend in CORE is much larger than the trend in CCSM4. Taken together, all of these differences indicate a much larger negative feedback in CORE of the heat flux component trends onto the SST warming trend than is evident in CCSM4.

To quantify the contrasting trend behavior, we present the Δ_{25} values for each heat flux component in Table 3 over the period from 1981 to the end of the CCSM4 twentieth-century integrations in 2005. As suggested by the differences in Fig. 8, there is a negative feedback $\Delta_{25}Q_{\rm as}$ of −11.3 W m^{-2} in CORE, while the CCSM4 feedback is slightly positive (1.2 W m^{-2}) giving a difference between CCSM4 and CORE of 12.5 W m^{-2}. Breaking the total air–sea heat flux $Q_{\rm as}$ into its two components [see Eq. (6)], we note that $\Delta_{25}Q_{\rm sky}$ and $\Delta_{25}Q_{\rm pbl}$ contribute to the overall 12.5 W m^{-2} difference almost equally with CCSM4–CORE differences of 6.1 and 6.4 W m^{-2}, respectively.

Focusing first on $Q_{\rm sky}$ [Eq. (1)] in Table 3, the largest contributor to the 6.1 W m^{-2} difference in $\Delta_{25}Q_{\rm sky}$ is

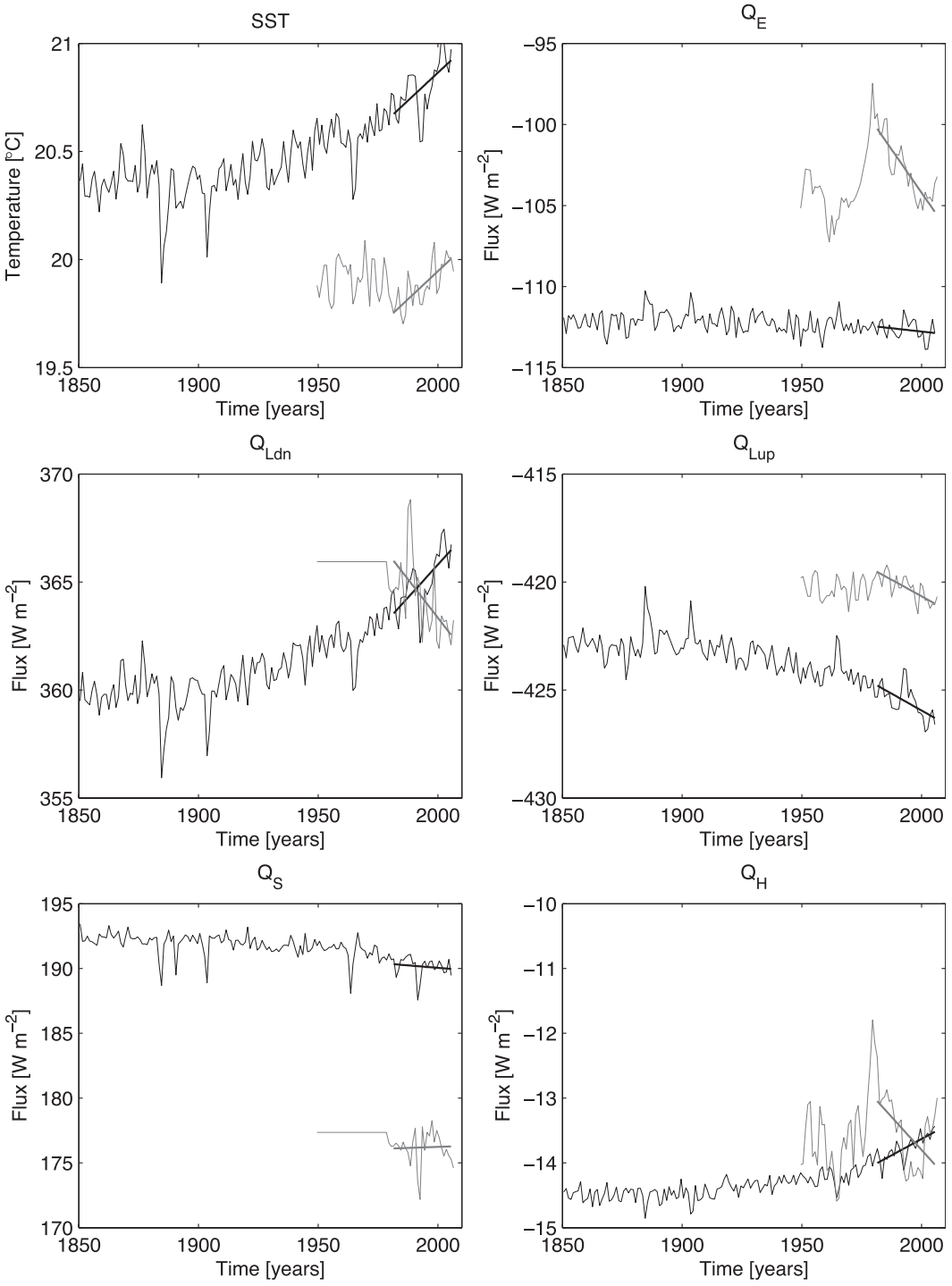

FIG. 8. Time series of globally averaged (over ice-free ocean points) SST and heat flux components of the one member from the CCSM4 20C ensemble. The time series spanning 1949–2005 is the CORE flux and Hurrell et al. (2008) SST data. The dark lines are a linear fit to the data from years 1981 to 2005. The increment in latitude is 15°.

$\Delta_{25}Q_{\mathrm{Ldn}}$ with little contribution from $\Delta_{25}Q_S$. Note that $\Delta_{25}Q_{\mathrm{Ldn}}$ is of opposite sign in CCSM4 compared to CORE, evident in Fig. 8, and contributes 6.6 W m^{-2} to the $\Delta_{25}Q_{\mathrm{sky}}$ difference. Also, $\Delta_{25}Q_S$ is of opposite sign in CCSM4 compared to CORE but, because it is small, does not contribute significantly to the overall $\Delta_{25}Q_{\mathrm{sky}}$ difference.

The negative trend of CORE Q_{Ldn} data is somewhat surprising given the increase in GHG concentrations

TABLE 3. Trends in air–sea heat fluxes over the ice-free ocean expressed as changes Δ_{25} over the 25 yr (1981–2005) in W m^{-2} from the linear regressions of Fig. 8 for both CCSM4 and CORE. The rightmost column gives the differences (CCSM4 − CORE). An asterisk denotes a component calculated in the 40°S–40°N band using Eqs. (10)–(13) as described in the text. The corresponding 25-yr change in SST for both CCSM4 and CORE is $\Delta_{25} = 0.26$°C. The components of a particular heat flux may not sum exactly due to rounding off and different geographical regions.

Heat flux component	CCSM4 Δ_{25}SST = 0.26°C	CORE Δ_{25}SST = 0.26°C	Difference
$\Delta_{25}Q_{as}$	1.2	−11.3	12.5
$\Delta_{25}Q_{sky}$	2.7	−3.4	6.1
$\Delta_{25}Q_{Ldn}$	3.1	−3.6	6.6
$\Delta_{25}Q_S$	−0.4	0.2	−0.5
$\Delta_{25}Q_{pbl}$	−1.5	−7.9	6.4
$\Delta_{25}Q_{Lup}$	−1.6	−1.5	−0.1
$\Delta_{25}Q_E$	−0.4	−5.3	4.9
Wind effect*	−0.2	−2.7	2.6
Δq effect*	−1.8	−3.0	1.2
$\Delta_{25}Q_H$	0.5	−1.0	1.5
Wind effect*	−0.02	−0.3	0.3
ΔT effect*	0.3	−1.3	1.6

over this time period. This issue was discussed in detail in Large and Yeager (2012) and is further investigated here. Over the ocean, CORE Q_{Ldn} is simply the uncorrected International Satellite Cloud Climatology Project (ISCCP-FD) product (Zhang et al. 2004), daily averaged and regridded. Figure 9 shows that, globally, Q_{Ldn} is in fact increasing because of a larger increase over land, which is more than enough to compensate the observed decrease over the ice-free ocean. Note that the increase in both the global and land-only Q_{Ldn} does not begin until approximately 1999; however, the decrease over the ocean has been occurring since 1984.

Turning our attention now to Q_{pbl} [Eq. (2)] in Table 3, we find that its Δ_{25} difference of 6.4 W m^{-2} is dominated by Q_E with a significant contribution from Q_H and little contribution from Q_{Lup}. The small difference in Q_{Lup} is consistent with the similar trends in SST in both datasets [see Eq. (3)]. As expected from Eq. (8), the $\Delta_{25}Q_E$ acts to diminish SST warming in both CORE and CCSM4, but much less so in CCSM4. Also, $\Delta_{25}Q_E$ is expected to decrease the SST warming trend and does so in the CORE data, but in CCSM4 the trend in Q_H is positive and therefore contributes to the SST warming. A further breakdown of the Q_E and Q_H PBL flux trends is discussed in the next section.

Using these Δ_{25} values, we estimate the feedback coefficient C_{pbl} and average over regions (Table 5). The various geographical areas are formed by combining regions shown in Fig. 4, so that each is large enough to neglect ocean advective transport effects on SST (Large and Yeager 2012). Excluded regions include marginal ice zones (regions 1, 2, 11, 12, 15, 19, and 20) and lower southern latitude regions (8–10) where advective transports of the Antarctic Circumpolar Current and the Agulhas Retroflection could be significant.

There is a stronger negative feedback due to Q_{pbl} in CORE than in CCSM4 (Table 5), with a global, ice-free ocean C_{pbl} twice as large in CORE as in CCSM4, indicating a much weaker negative feedback of the PBL fluxes on SST in CCSM4. As expected, C_{pbl} is negative everywhere in both datasets, but it varies by region and is much less negative everywhere in the CCSM4 simulation compared to CORE. Regions of particularly underestimated feedback are the Indian Ocean (by a factor of 5), the equatorial Pacific, and the South Pacific.

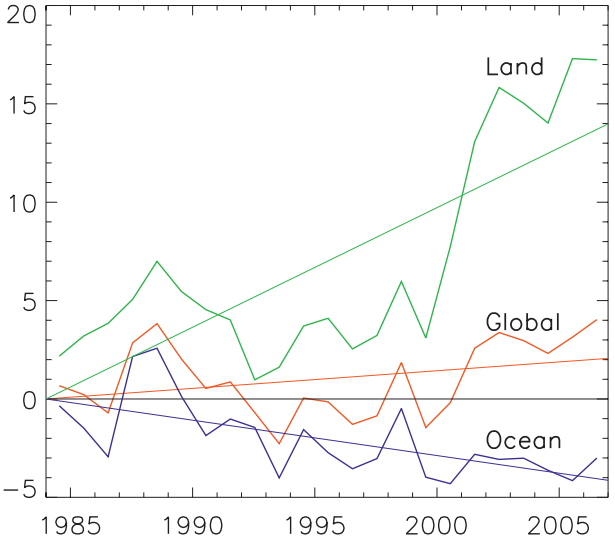

FIG. 9. Downwelling longwave radiation at the surface from ISSCP-FD, annually averaged over land (green), the ice-free ocean (blue), and the globe (red) from 1984 through 2006. For reference, the respective 23-yr averages are 317, 364, and 344 W m^{-2}. Linear regressions of each time series give the straight line trends shown from the beginning of 1984 to the end of 2006. Values plotted are the differences from the regression values at the beginning of 1984.

The global ice-free C_{pbl} can be compared to a similar coefficient calculated from the numbers in Table 3, although the results will be slightly different (because of the order in which spatial averaging and regression are performed). In Table 3, spatial averages of the flux components are first computed and then Δ_{25} values found from the averaged time series, whereas C_{pbl} is computed from spatial averages of the Δ_{25} values. Regardless of the method, CCSM4 shows a much weaker PBL heat flux feedback than CORE.

c. Sensible and latent heat fluxes

Why does the behavior of the latent and sensible heat fluxes, and hence Q_{pbl} and C_{pbl}, differ so much between CCSM4 and CORE? The answer to this question has implications for verifying the model's physics and feedbacks. From the previous section, we find that the Q_E and Q_H flux components largely explain the $\Delta_{25}Q_{\mathrm{pbl}}$ difference between CCSM4 and CORE. Their respective contributions to the 6.4 W m^{-2} $\Delta_{25}Q_{\mathrm{pbl}}$ difference are 4.9 and 1.5 W m^{-2} (Table 3). From Eqs. (4) and (5), the difference can be further apportioned between the 25-yr changes in Δq (the "Δq effect" on Q_E, denoted $\delta_{25}^{\Delta q}Q_E$), ΔT (the "ΔT effect" on Q_H, denoted $\delta_{25}^{\Delta T}Q_H$), and U (the "wind effect" on Q_E and on Q_H, denoted $\delta_{25}^{U}Q_E$ and $\delta_{25}^{U}Q_H$, respectively). We estimate these effects within both CCSM4 and CORE by substituting "typical values" and Δ_{25} values for each variable into partial derivatives of Eqs. (4) and (5) as follows:

$$\delta_{25}^{U}Q_E = (Q_E)\frac{\Delta_{25}U}{U}, \tag{10}$$

$$\delta_{25}^{U}Q_H = (Q_H)\frac{\Delta_{25}U}{U}, \tag{11}$$

$$\delta_{25}^{\Delta q}Q_E = (Q_E)\frac{\Delta_{25}\Delta q}{\Delta q}, \tag{12}$$

$$\delta_{25}^{\Delta T}Q_H = (Q_H)\frac{\Delta_{25}\Delta T}{\Delta T}, \tag{13}$$

where the δ_{25} denotes less rigorous (e.g., we used wind speed U and not $|\mathbf{U}|$) and consistent (e.g., all terms, except the Δ_{25} terms, are estimates) calculations of the 25-yr changes than the Δ_{25} values derived formally from linear regressions. We restrict this analysis to the latitudinal band 40°S–40°N in order to exclude any influence of changes in sea ice coverage, especially in light of the large decrease in Arctic sea ice cover in the late twentieth century in both the model and observations, and because the PBL flux has its largest response to SST in this band (Large and Yeager 2012). The "typical" values and resulting 25-yr changes for both CCSM4 and CORE are given in Table 4. Typical values of Q_E, Q_H, U, Δq, and ΔT are taken from 25-yr time series averaged over the band (not shown) with Δ_{25} values of the latter three calculated from linear regressions of these time series. The Δ_{25} values for ΔT and Δq can also be obtained from Fig. 10. The resulting contributions of these "effect" terms are inserted into Table 3 as rough estimates, but they are not equivalent to the other values in the table that are calculated over the ice-free ocean. Consequently, the contribution of the wind, ΔT, and Δq effects will not sum to equal the $\Delta_{25}Q_{\mathrm{pbl}}$ values.

TABLE 4. Typical values and 25-yr changes of the variables in Eqs. (10)–(12) averaged over the 40°S–40°N band. See text for description.

	CCSM4	CORE	Units
Q_E	−131	−123	W m^{-2}
Q_H	−12.7	−13.5	W m^{-2}
U	8.0	5.0	m s^{-1}
Δq	−8.8	−8.8	g kg^{-1}
ΔT	−1.2	−1.3	°C
$\Delta_{25}U$	0.01	0.1	m s^{-1}
$\Delta_{25}\Delta q$	−0.1	−0.2	g kg^{-1}
$\Delta_{25}\Delta T$	0.03	−0.1	°C
$\delta_{25}^{U}Q_E$	−0.2	−2.7	W m^{-2}
$\delta_{25}^{U}Q_H$	−0.02	−0.3	W m^{-2}
$\delta_{25}^{\Delta q}Q_E$	−1.8	−3.0	W m^{-2}
$\delta_{25}^{\Delta T}Q_H$	0.3	−1.3	W m^{-2}

From the band averaged numbers, we are able to attribute $\Delta_{25}Q_{\mathrm{pbl}}$ CCSM4–CORE differences to the wind, ΔT, and Δq effects (Table 3). The largest contributor is the difference in the wind effect on Q_E. The ΔT and Δq effects have similar contributions, while the wind effect on Q_H plays the least role in CCSM4–CORE differences of $\Delta_{25}Q_{\mathrm{pbl}}$. The relative contribution of these terms to the overall $\Delta_{25}Q_{\mathrm{pbl}}$ CCSM4–CORE difference is apportioned as follows: 45% is due to the wind effect on Q_E, 29% to the ΔT effect on Q_H, 21% to the Δq effect on Q_E, and 5% to the wind effect on Q_H. We also calculated these effects using global, ice-free points (not shown); although individual contributions of these effects to Q_E and Q_H in CCSM4 and CORE can be somewhat different, the CCSM4–CORE differences are similarly apportioned.

The differences between ΔT and Δq behavior between CCSM4 and CORE are attributed to the differences in the T_{abl} and q_{abl} trends because we know that the SST and SSq trends are very similar between the two. Indeed, from Fig. 10, both T_{abl} and q_{abl} increase faster in CCSM4 than CORE. The value of T_{abl} rises faster than SST such that the trend in ΔT is slightly positive and of opposite sign from the ΔT trend in CORE. In CCSM4 the accompanying increase in ABL humidity q_{abl} is initially about the same as SSq (little change in Δq) but then

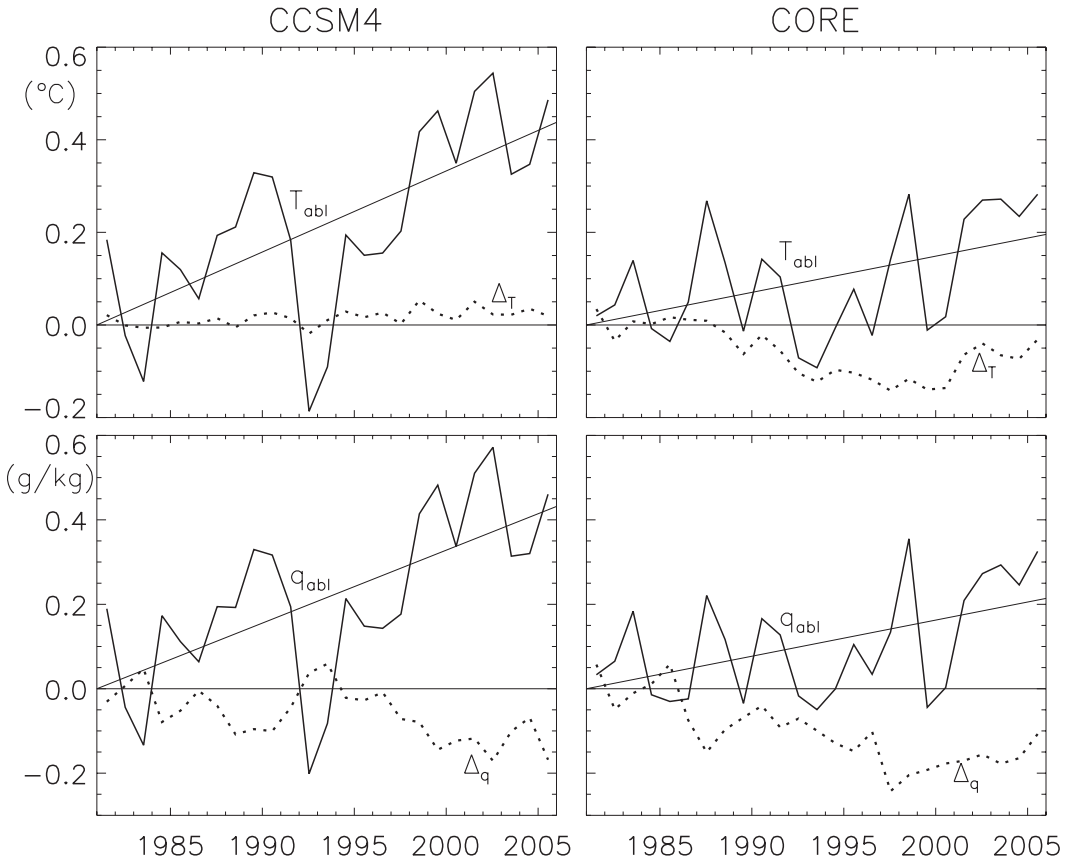

FIG. 10. Time series of the (top) air–sea temperature and (bottom) humidity differences relative to their 1980 value averaged over 40°S to 40°N for (left) CCSM4 and (right) CORE. The dotted lines are the air–sea temperature and humidity differences (ΔT defined as $T_{\rm abl}$ − SST and Δq defined as $q_{\rm abl}$ − SSq) and the solid lines are the 10-m air temperature ($T_{\rm abl}$) and humidity ($q_{\rm abl}$). The straight lines in each plot are the regression of the $T_{\rm abl}$ and $q_{\rm abl}$ time series.

increases at a slower rate such that Δq remains below its 1981 value only after 1994. In CORE, however, Δq drops below its 1981 value much earlier, starting in 1986. The relatively greater warming and moistening of the marine ABL in CCSM4 is an important difference from CORE because of the effects on the sensible and latent heat fluxes (Table 3) and on the air–sea coupling (Table 5).

The increases in CORE warming and moistening may be due to shifts in natural modes of climate variability, such as the Pacific decadal oscillation and the North Atlantic Oscillation, that CCSM4 would not be expected to represent coincidently. Similarly, at higher frequencies CCSM4 and CORE variability should not, and do not, track each other, but ABL temperature and moisture are highly correlated in both. Similar high correlations with SST are implied by the relatively muted and uncorrelated variability of Δq and ΔT.

The conclusion from this section is that the CCSM4 fluxes do not respond to an increase in SST in the same manner as the CORE fluxes over the ocean, which may lead to differences in heating trends. Different behaviors in $Q_{\rm Ldn}$ over land and ocean contribute to the overall flux difference, but the most impactful differences are seen in the latent and sensible heat fluxes of the PBL. Because the ABL warms and moistens faster in the model than CORE, the negative feedback of Q_E and

TABLE 5. The value of $C_{\rm pbl}$ [Eq. (9)] averaged over geographical regions. Units: W m^{-2} °C^{-1}.

Geographical region	Region numbers	CORE $C_{\rm pbl}$	CCSM4 $C_{\rm pbl}$	Ratio
Global	1–21	−28	−14	2
North Atlantic	3, 4, 5	−18	−7	2.6
Equatorial Atlantic	6, 7	−29	−14	2.1
Indian	13, 14	−46	−9	5.1
North Pacific	16, 17	−18	−6	3
Equatorial Pacific	21	−35	−8	4.4
South Pacific	18	−33	−9	3.7

Q_H is either nonexistent or much less than that found in CORE in relation to a positively trending SST.

6. Discussion and conclusions

Eliminating the mean bias in climate models continues to be a principal motivating goal in climate science, and thus, we investigated mean biases in air–sea heat and freshwater flux components in the CCSM4. Although regional biases in some components are improved, others are still present and some are made worse. The largest degradation in the transition from CCSM3 to CCSM4 in mean flux bias is found in the greater shortwave radiation reaching the ocean's surface. This degradation is a global, nearly uniform increase with many regional averages falling outside the range of observation-based estimates.

Neale et al. (2012, manuscript submitted to *J. Climate*) document that cloud fraction is reduced globally because of the addition of new parameterizations in CCSM4, and results from this study support this conclusion. In the regional time mean, some compensation to excess shortwave radiation in CCSM4 is provided by a deficit in net longwave radiation. Additionally, the error variance of downward longwave radiation shares similar patterns with the error variance of shortwave radiation. Too little cloud cover would allow more shortwave radiation to reach the ocean's surface and lead to less downward longwave radiation due to less absorption of upward longwave radiation.

Although downward longwave radiation provides some compensation to excess shortwave radiation in the mean, the majority of compensation occurs with the latent heat flux. The excess shortwave radiation heats the surface, warming SST until evaporation compensates. This balance between heat flux terms keeps the net air–sea heat flux close to CORE and within observation-based estimates, implying reasonable ocean heat transports. Without similar compensation among freshwater flux components, the enhanced evaporation leads to net air–sea freshwater fluxes that can be outside the range of observation-based estimates, and thus lead to erroneous ocean salinity and density. Enhanced evaporation can also lead to an enhanced hydrological cycle with more precipitation over both the ocean and land.

The variability of the air–sea flux fields could be significantly improved. The need for improvement is not inconsequential for understanding climate change, as the seasonal cycle and interannual variability are strongly correlated with societal and climatological interest (Stine et al. 2009). Annual variability is substantially in error in virtually all regions with the likelihood of robust CCSM4–CORE disagreement, based on the WPA analysis of Stevenson et al. (2010), almost always above 90%. Using WPA, representation of interannual variability is better with all of the regions showing likely agreement. Geographic comparisons of the variances reveal that the model often agrees with CORE data in the basic geographical pattern of variance for each variable, with regions of higher variance in CORE also having higher variance in CCSM4 (not shown). The standard deviation of error is typically smaller than the mean bias. However, the variance errors perhaps belie many misrepresented processes. As annual and interannual variability in fluxes have not been examined in previous version of this model, this assessment may also serve as a variability benchmark for future model versions.

The regional patterns of interannual, annual, and mean bias often differ for a given air–sea flux type. By analyzing these differences, the processes still needing improvement may be isolated not only by region but by time scale. As an example, the net shortwave radiation has the largest errors on all time scales (mean, annual, and interannual); however, the pattern of errors is different for each time scale, suggesting that cloud activity at each time scale may be flawed with different patterns. Mean shortwave differences are largely over the eastern subtropical basins where stratus cloud decks are often poorly modeled (Zheng et al. 2011), while annual and interannual variability error patterns indicate a relationship with ITCZ and equatorial cloud errors, respectively. The significant errors on annual time scales should draw attention to the need for better modeling of the air–sea processes that dominate these time scales, such as exchange between the PBLs and ocean interior and free atmosphere, dynamics affecting the seasonal cycle of the oceanic mixed layer, monsoons, and seasonal cloud variability. Fortunately, observations spanning the annual cycle are common, and simulations to test sensitivity of the annual cycle are less expensive than those for the interannual variability. Thus, substantial improvements in annual variance fidelity may be possible in future model versions.

Differences in sensible and latent heat flux trends explain approximately half of the late twentieth-century trend difference in air–sea heat flux between CCSM4 and CORE, with a large portion of these differences due to differences in ABL properties. If the true fluxes constitute a negative feedback to SST anomalies, we expect that an increase in SST would lead to a decrease of heat flux into the ocean and thus impede further SST warming. While the 25-yr trends indicate that this is happening in the CORE data, it is not occurring in any of the CCSM4 20C ensemble members investigated here. In fact, the sensible heat flux trend continues to add more and more heat to the ocean over the 25 years

counter to the observation-based estimates we have. Similarly, the latent heat flux trend in the 20C simulation, though of the same sign as in CORE, is of much smaller magnitude and therefore is not removing as much heat from the ocean over these 25 years as in CORE. The reason for these two discrepancies lies in the fact that ABL temperature and humidity in the 20C simulation track sea surface change too closely and therefore allow for more heating of the sea surface.

There are a number of ways the marine ABL might warm and moisten in response to greater GHG concentrations. First, a warmer SST, possibly due to more Q_{Ldn}, would transfer some of the ocean's excess heat and effectively infinite water to the ABL. CORE behavior is more consistent with this mechanism than CCSM4, whose marine atmosphere warms more than SST, such that there is less heat transferred from the ocean's surface to the ABL. However, ocean processes are likely to be involved in the observed SST rise in CORE (Large and Yeager 2012). Second, the properties of the marine ABL depend on mixing with the free troposphere, which is a longstanding challenge for climate models. This mechanism alone would not explain warming and moistening since more mixing with the troposphere would likely warm and dry the ABL. Third, a warmer troposphere, including the ABL, could then hold more water vapor, and therefore absorb more Q_{Lup}, as well as downwelling longwave radiation from clouds, meaning that the ABL would warm faster than otherwise. The latter is the opacity effect of Stephens et al. (2012), which they state would reduce Q_{Ldn}, and hence warming of SST, through more absorption and less emission of Q_{Ldn}. Such a scenario is consistent with changes in the marine ABL of CCSM4. Such a direct radiative response may vary seasonally and regionally. Differing behavior between land and ocean, in particular, presents a fourth possibility of greater ABL warming over land followed by advection of this air over the ocean. Further investigation is needed to determine the mechanisms at work in both CORE and CCSM4.

Large and Yeager (2012) suggest that ocean processes in response to internal natural variability may be responsible for the late twentieth-century SST changes. If natural variability is a factor, we do not expect the model to represent the SST rise in the same way since its variability will not be synchronous with nature. Time will reveal the behavior of SST in nature and whether it continues on the same upward trend or if negative feedbacks associated with the PBL fluxes will slow the SST increase. How other models' ABLs behave is outside the scope of this work, but we propose the use of the C_{pbl} term [Eq. (9)] for assessment of any model's boundary layer scheme.

Acknowledgments. The CESM project is supported by the National Science Foundation and the Office of Science (BER) of the U.S. Department of Energy. S. Stevenson was supported by NASA Grant NNX09A020H and B. Fox-Kemper by Grants NSF 0934737 and NASA NNX09AF38G. Computing resources were provided by the Climate Simulation Laboratory at NCAR's Computational and Information Systems Laboratory (CISL), sponsored by the National Science Foundation and other agencies. This research was enabled by CISL compute and storage resources. Bluefire, a 4064-processor IBM Power6 resource with a peak of 77 TeraFLOPS provided more than 7.5 million computing hours, the GLADE high-speed disk resources provided 0.4 petabytes of dedicated disk and CISL's 12-PB HPSS archive provided over 1 petabyte of storage in support of this research project.

REFERENCES

Bryan, F. O., B. G. Kauffman, W. G. Large, and P. R. Gent, 1996: The NCAR CSM flux coupler. NCAR Tech. Rep. NCAR/TN-424-STR, 53 pp.

Collins, W. D., and Coauthors, 2006: The Community Climate System Model version 3 (CCSM3). *J. Climate,* **19,** 2122–2143.

Dai, A., T. Qian, K. E. Trenberth, and J. D. Milliman, 2009: Changes in continental freshwater discharge from 1948 to 2004. *J. Climate,* **22,** 2773–2792.

Danabasoglu, G., S. C. Bates, B. P. Briegleb, S. R. Jayne, M. Jochum, W. G. Large, S. Peacock, and S. G. Yeager, 2011: The CCSM4 ocean component. *J. Climate,* **25,** 1361–1389.

Deser, C., and Coauthors, 2012: ENSO and Pacific decadal variability in Community Climate System Model version 4. *J. Climate,* **25,** 2622–2651.

Fox-Kemper, B., and Coauthors, 2011: Parameterization of mixed layer eddies. III: Implementation and impact in global ocean climate simulations. *Ocean Modell.,* **39,** 61–78.

Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate,* **24,** 4973–4991.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.,* **113,** D06104, doi:10.1029/2007JD008972.

Griffies, S. M., and Coauthors, 2009: Coordinated Ocean-Ice Reference Experiments (COREs). *Ocean Modell.,* **26,** 1–46.

Hack, J. J., J. M. Caron, S. G. Yeager, K. W. Oleson, M. M. Holland, J. E. Truesdale, and P. J. Rasch, 2006: Simulation of the global hydrological cycle in the CCSM Community Atmosphere Model version 3 (CAM3): Mean features. *J. Climate,* **19,** 2199–2221.

Holland, M. M., C. M. Bitz, E. C. Hunke, W. H. Lipscomb, and J. L. Schramm, 2006: Influence of sea ice thickness distribution of polar climate in CCSM3. *J. Climate,* **19,** 2398–2414.

Hurrell, J., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Climate,* **21,** 5145–5153.

Jahn, A., and Coauthors, 2012: Late-twentieth-century simulation of Arctic sea ice and ocean properties in the CCSM4. *J. Climate,* **25,** 1431–1452.

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

Kaspi, Y., and T. Schneider, 2011: Winter cold of eastern continental boundaries induced by warm ocean waters. *Nature,* **471,** 621–624, doi:10.1038/nature09924.

Large, W. G., and G. Danabasoglu, 2006: Attribution and impacts of upper-ocean biases in CCSM3. *J. Climate,* **19,** 2325–2346.

——, and S. Yeager, 2009: The global climatology of an interannually varying air–sea flux dataset. *Climate Dyn.,* **33,** 341–364, doi:10.1007/s00382-008-0441-3.

——, and S. G. Yeager, 2012: On the observed trends and changes in global sea surface temperature and air–sea heat fluxes (1984–2006). *J. Climate,* **25,** 6123–6135.

McClean, J. L., and Coauthors, 2012: A prototype two-decade fully-coupled fine-resolution CCSM simulation. *Ocean Modell.,* **39,** 10–30, doi:10.1016/j.ocemod.2011.02.011.

Neale, R., J. Richter, and M. Jochum, 2008: The impact of convection on ENSO: From a delayed oscillator to a series of events. *J. Climate,* **21,** 5904–5924.

Richter, J., and P. Rasch, 2008: Effects of convective momentum transport on the atmospheric circulation in the Community Atmosphere Model, version 3. *J. Climate,* **21,** 1487–1499.

Roske, F., 2006: A global heat and freshwater forcing dataset for ocean models. *Ocean Modell.,* **11,** 235–297.

Smith, R., and Coauthors, 2010: The Parallel Ocean Program (POP) reference manual. Los Alamos National Laboratory Tech. Rep. LAUR-10-01853, 140 pp.

Stephens, G. L., M. Wild, P. W. Stackhouse Jr., T. L'Ecuyer, S. Kato, and D. S. Henderson, 2012: The global character of the flux of downward longwave radiation. *J. Climate,* **25,** 2329–2340.

Stevenson, S., B. Fox-Kemper, M. Jochum, B. Rajagopalan, and S. G. Yeager, 2010: ENSO model validation using wavelet probability analysis. *J. Climate,* **23,** 5540–5547.

——, ——, ——, R. Neale, C. Deser, and G. Meehl, 2012: Will there be a significant change to El Niño in the twenty-first century? *J. Climate,* **25,** 2129–2145.

Stine, A., P. Huybers, and I. Fung, 2009: Changes in the phase of the annual cycle of surface temperature. *Nature,* **457,** 435–440.

Taylor, P. K., 2000: Final report of the joint WCRP/SCOR working group on air–sea fluxes: Intercomparison and validation of ocean–atmosphere energy flux fields. WCRP-112, WMO/TD-1036, 303 pp.

Torrence, C., and G. P. Compo, 1998: A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.,* **79,** 61–78.

Wittenberg, A., A. Rosati, N. Lau, and J. Ploshay, 2006: GFDL's CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. *J. Climate,* **19,** 698–722.

Xie, P., and P. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.,* **78,** 2539–2558.

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

Zheng, Y., T. Shinoda, J.-L. Lin, and G. N. Kiladis, 2011: Sea surface temperature biases under the stratus cloud deck in the southeast Pacific Ocean in 19 IPCC AR4 coupled general circulation models. *J. Climate,* **24,** 4139–4164.