Evaluation of Surface Flux Parameterizations with Long-Term ARM Observations

GANG LIU

Brookhaven National Laboratory, Upton, New York, and School of Atmospheric Sciences, Nanjing University, Nanjing, China

YANGANG LIU AND SATOSHI ENDO

Brookhaven National Laboratory, Upton, New York

(Manuscript received 28 March 2012, in final form 6 September 2012)

ABSTRACT

Surface momentum, sensible heat, and latent heat fluxes are critical for atmospheric processes such as clouds and precipitation, and are parameterized in a variety of models ranging from cloud-resolving models to large-scale weather and climate models. However, direct evaluation of the parameterization schemes for these surface fluxes is rare due to limited observations. This study takes advantage of the long-term observations of surface fluxes collected at the Southern Great Plains site by the Department of Energy Atmospheric Radiation Measurement program to evaluate the six surface flux parameterization schemes commonly used in the Weather Research and Forecasting (WRF) model and three U.S. general circulation models (GCMs). The unprecedented 7-yr-long measurements by the eddy correlation (EC) and energy balance Bowen ratio (EBBR) methods permit statistical evaluation of all six parameterizations under a variety of stability conditions, diurnal cycles, and seasonal variations. The statistical analyses show that the momentum flux parameterization agrees best with the EC observations, followed by latent heat flux, sensible heat flux, and evaporation ratio/Bowen ratio. The overall performance of the parameterizations depends on atmospheric stability, being best under neutral stratification and deteriorating toward both more stable and more unstable conditions. Further diagnostic analysis reveals that in addition to the parameterization schemes themselves, the discrepancies between observed and parameterized sensible and latent heat fluxes may stem from inadequate use of input variables such as surface temperature, moisture availability, and roughness length. The results demonstrate the need for improving the land surface models and measurements of surface properties, which would permit the evaluation of full land surface models.

1. Introduction

Surface momentum, sensible heat, and latent heat fluxes are critical for atmospheric processes such as clouds and precipitation, and are often parameterized in a variety of models due to limited grid resolution in these models, such as the Weather Research and Forecasting (WRF) model (Skamarock et al. 2008) and general circulation models (GCMs). In numerical models, these turbulent flux parameterizations are collectively referred to as the surface flux parameterization (SFP), and through SFP, the atmosphere is coupled with the underlying surface.

Evaluation of SFP schemes is essential to any model development. There are generally two approaches for evaluating parameterizations: direct offline evaluation and full-model online evaluation. Many studies on the SFP schemes have been conducted in an ''online'' mode (e.g., Betts et al. 1997; Chen et al. 1997), whereby numerical models are run with different SFP schemes, and the impact of the SFP schemes on the simulation results of the numerical models are evaluated against observations. With the online evaluation, the impact of the SFP schemes on the corresponding models can be investigated. The SFP schemes, however, themselves cannot be evaluated unambiguously with the online mode, since the parameterized turbulent fluxes are related to resolved meteorological quantities (e.g., wind speed, air temperature, humidity, and ground temperature), which are predicted by the numerical model rather than observed. The errors in the model-predicted quantities will in turn lead to errors in the parameterized turbulent fluxes.

The direct offline evaluation of the SFP schemes minimizes the compound errors associated with the full

Corresponding author address: Gang Liu, School of Atmospheric Sciences, Nanjing University, Nanjing 210093, China.
E-mail: gangliu@nju.edu.cn

DOI: 10.1175/MWR-D-12-00095.1

model evaluation, and is the focus of this paper. In the offline mode, the turbulent fluxes are calculated by the SFP schemes using the corresponding measurements of mean meteorological quantities as inputs, and the parameterized turbulent fluxes are evaluated against the concurrent measurements of surface turbulent fluxes. Direct offline SFP evaluation against observations is limited, due to the lack of long-term and continuous flux observations in the surface layer. This has hindered proper assessment of the SFP schemes and understanding of turbulent transfer between the atmosphere and the surface. The rare study by Cassano et al. (2001) evaluated seven SFP schemes, but was limited in several aspects by the scarcity of observations. Only 45 months' worth of data collected in Antarctica under stable stratification conditions with a temporal resolution of 1 h were used, and there were no comparisons for the latent heat flux.

The U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement (ARM) program (www.arm.gov) has conducted continuous measurements of surface turbulent fluxes at the Southern Great Plains (SGP) site, by use of energy balance Bowen ratio (EBBR) stations since 1993, and by the use of eddy correlation (EC) stations since 1997. The EC method provides measurements of momentum and sensible heat and latent heat fluxes, while the EBBR method only yields results for the latter two. This study takes advantage of these long-term observations to evaluate the SFP schemes commonly used in the WRF model and in three major U.S. GCMs that participate in the Fast-Physics System Testbed and Research (FASTER) project (www.bnl.gov/esm): the Goddard Institute for Space Studies (GISS) GCM (Schmidt et al. 2006), the Geophysical Fluid Dynamics Laboratory (GFDL) GCM [global atmosphere and land model; GAMDT (2004)], and the National Center for Atmospheric Research (NCAR) Community Atmosphere Model (CAM; Collins et al. 2004).

The rest of the paper is organized as follows. The ARM SGP observations related to this study are described in section 2, followed by the description of the evaluated SFP schemes in section 3. The results from the six SFP schemes and comparisons with the observed surface fluxes are presented in section 4. The possible factors for poorly parameterized sensible and latent heat fluxes are discussed in section 5. This study is summarized in section 6.

2. Measurements related to this study

A primary objective of the ARM program is to improve scientific understanding of the fundamental physics related to interactions between clouds and radiative processes in the atmosphere, with emphasis on making continuous field measurements that enhance the evaluation and parameterization of cloud-related fast processes in climate models (Stokes and Schwartz 1994; Ackerman and Stokes 2003). The SGP site was the first field measurement site established by the ARM program, and the SGP Central Facility (CF) located near Lamont in north-central Oklahoma (36°36′18.0″N, 97°29′6.0″W, and 320 m above sea level) houses the core instruments. The central facility site is selected for this study since all of the required coincident EC, EBBR, and infrared thermometer (IRT) measurements were available there; these measurements are described in the following.

The EC measurement system at 3-m height consists of a fast-response sonic anemometer for measuring three-dimensional winds and the speed of sound used to derive the air temperature, as well as an open-path infrared gas analyzer for the water vapor density and the CO_2 concentration. The eddy covariance technique is applied to the original measurements to derive 30-min surface turbulent fluxes of momentum, sensible heat, latent heat, and carbon dioxide. The EBBR flux measurement system produces 30-min estimates of the vertical fluxes of sensible and latent heat from measurements of net radiation, soil surface heat flux, and the vertical gradients of temperature and relative humidity.

According to previous studies (Twine et al. 2000; Brotzge and Crawford 2003; Jiang et al. 2004), the typical error in EC measurements of the sensible and latent heat fluxes is on the order of [~(10%–30%)], while the error in EBBR measurements is typically on the order of [~(10%–20%)]. Generally, the errors are at a minimum during the late morning and afternoon when the atmospheric boundary layer is unstable with strong mixing, while they are at a maximum during the early morning and night when the atmospheric boundary layer is stable or at transitional stages.

Direct offline evaluation also needs the ground temperature as an input to the SFP schemes. At SGP, the ground temperature is measured with a downward-pointing IRT located at 10-m height above ground level. The IRT is a radiation pyrometer that measures the equivalent blackbody brightness temperature of the scene in its field of view, and provides a skin temperature every minute. We aggregate the 1-min IRT measurements into the 30-min averages to be consistent with the temporal resolution of the EC and EBBR flux measurements.

The period that has all the required coincident EC, EBBR, and IRT measurements spans from 0000 UTC 12 September 2003 to 2330 UTC 13 August 2010. These

~7 yr of data are used in this study. The ARM SGP measurements were used in some previous evaluations (see the special issue of *Mon. Wea. Rev.*, 2006, Vol. 134, No. 1; Yang et al. 2006), but the length of the dataset used in those studies was limited, ranging from a few months to a few years. The 7-yr dataset used here is the longest that has been analyzed so far, providing unprecedented statistics under a wide range of stability conditions. This is unique compared to previous studies.

The ARM SGP site is representative of a continental climate in the midlatitudes. Although the 7-yr dataset is not long enough to be very representative of the region or for detecting interannual variations, the period is the longest available to us that has all the required coincident EC, EBBR, and IRT measurements. In the original 7-yr data record at the ARM SGP CF site, several gaps appear due to sensor problems with the EBBR and EC, and some data were marked as questionable when the EC sonic anemometer was not operating properly under wet conditions. To assure data quality, the gaps and flux data collected during the days with rainfall/snow are not analyzed in this study.

3. Description of the surface flux parameterization schemes

a. General description

The sensible heat flux (SHF), latent heat flux (LHF), and momentum flux (MF) are defined as

$$\mathrm{SHF} = \rho C_p \overline{w'\theta'}, \tag{1}$$

$$\mathrm{LHF} = \rho L_v \overline{w'q'}, \quad \text{and} \tag{2}$$

$$\mathrm{MF} = -\rho \overline{u'w'} = \rho u_*^2, \tag{3}$$

where ρ is the density of air; C_p is the specific heat of air at constant pressure; L_v is the latent heat of the vaporization of water; θ is the potential temperature; q is the specific humidity; u and w are horizontal and vertical wind speeds, respectively; the prime denotes fluctuation from the average; $\overline{w'\theta'}$, $\overline{w'q'}$, and $-\overline{u'w'}$ are the kinematic definitions of the sensible heat flux, latent heat flux, and momentum flux, respectively; and u_* is the friction velocity. Note that Eq. (3) is obtained when the coordinate system is aligned so that the x axis points in the direction of the surface stress and the component $-\overline{vw'}$ is thus eliminated.

In atmospheric models, the standard approach for calculating a surface flux is expressing the surface flux as the difference in the corresponding mean quantity between the surface and the lowest model level assumed to be in the surface layer, and assuming the validity of the Monin–Obukhov similarity theory (MOST; Monin and Obukhov, 1954) in the surface layer. Mathematically, the equations in the kinematic forms are written as

$$\overline{w'\theta'} = C_h |U|(\theta_s - \theta), \tag{4}$$

$$\overline{w'q'} = C_q |U| A_m (q_s - q), \tag{5}$$

$$u_*^2 = C_m |U|^2, \tag{6}$$

$$C_h = k^2 / \left\{ \left[\ln\left(\frac{z}{z_0}\right) - \Psi_m\left(\frac{z}{L}\right) + \Psi_m\left(\frac{z_0}{L}\right) \right] \times \left[\ln\left(\frac{z}{z_0}\right) - \Psi_h\left(\frac{z}{L}\right) + \Psi_h\left(\frac{z_0}{L}\right) \right] \right\}, \tag{7}$$

$$C_q = C_h, \tag{8}$$

$$C_m = k^2 / \left[\ln\left(\frac{z}{z_0}\right) - \Psi_m\left(\frac{z}{L}\right) + \Psi_m\left(\frac{z_0}{L}\right) \right]^2, \tag{9}$$

$$|U|^2 = U^2 + V^2, \quad \text{and} \tag{10}$$

$$L = \frac{-u_*^3}{k \frac{g}{\overline{\theta_v}} \overline{w'\theta_v'}}. \tag{11}$$

The notation in the equations follows commonly used conventions. Briefly, $|U|$ is the mean speed of the wind vector at the height z; U and V are the large-scale horizontal velocity components; θ_s is the potential temperature at the surface; θ is the potential temperature at the height z; A_m is the moisture availability introduced as a measure of the degree of saturation at the ground (Zhang and Anthes 1982); q_s is the saturation specific humidity at the surface temperature; q is the specific humidity at the height z; $C_{h,q,m}$ are the transfer coefficients for heat, moisture, and momentum; k is the von Kármán constant (assumed to be 0.4); z is the reference height (i.e., in a weather–climate model, the height of the lowest model level; in experiments, the measurement height); z_0 is the roughness length (note that some schemes use different values for the momentum and heat–moisture roughness length); L is the Monin–Obukhov length; $g/\overline{\theta_v}$ is the buoyancy parameter; g is the gravity acceleration, taken to be 9.8 m s^{-2}; $\overline{\theta_v}$ is the mean virtual potential temperature; and $\overline{w'\theta_v'}$ is the virtual potential temperature flux.

The symbol $\Psi_{m,h}$ denote the stability profile function for momentum and heat (usually, the stability functions for heat and moisture are assumed to be the same).

TABLE 1. Major features and distinctions of the six SFP schemes evaluated.

SFP scheme	Outline of major features and distinctions
MM5	• The Beljaars (1994) correction is applied to calculate the convective velocity scale • Stability functions: Holtslag and De Bruin (1988) for stable conditions Paulson (1970) for unstable conditions
Eta	• The effects of the viscous sublayer are taken into account by introducing the roughness length for temperature and humidity (Zilitinkevich 1995) • The Beljaars (1994) correction is also applied to calculate the convective velocity scale under free convection conditions • The stability functions are based on Paulson (1970) for unstable conditions and Holtslag and De Bruin (1988) for stable conditions
PX	• Accounting for the difference in the sink–source heights of heat and momentum with parameterization of a viscous sublayer • The surface layer similarity functions are analytically estimated from large-scale state variables
CAM	• The roughness lengths for momentum and heat are based on Zilitinkevich (1970) • The scalar wind speed is defined as $\lvert U\rvert^2 = U^2 + V^2 + U_c^2,$ $U_c = \begin{cases} 0.1\ \mathrm{m\,s^{-1}}, & \text{if } \dfrac{z}{L} \geq 0 \text{ (stable)} \\ \beta w_* = \beta\left(z_i \dfrac{g}{T}\overline{w'\theta_v'}\right)^{1/3}, & \text{if } \dfrac{z}{L} < 0 \text{ (unstable)} \end{cases}$
GISS	• The roughness length for temperature over land is based on Brutsaert (1982) • The stability functions of Paulson (1970) and Holtslag and De Bruin (1988) are used, but with different constants in the functions
GFDL	• Similar to the GISS scheme, but with different empirical constants in the stability functions of Paulson (1970) and Holtslag and De Bruin (1988)

Various forms of the $\Psi_{m,h}$ function have been obtained (e.g., Businger et al. 1971; Dyer 1974; Högström 1988; Stull 1988). As the SFP schemes use the same MOST theoretical framework, the differences lie mainly in their specifications of the stability profile function and the various empirical parameters embedded in the different parameterizations.

In our offline evaluation, we use T, p, r, U, and V measured by the EC flux measurement system with the sensors located at 3 m above ground level (i.e., z is 3 m). Due to the lack of long-term data on detailed surface properties such as vegetation structure required as inputs to estimate the roughness length quantitatively, a value of 0.035 m is used for the roughness length in this study (Brown et al. 2002). The moisture availability A_m is obtained by looking up a classification chart in the version 3.0 WRF model with the known land type where the observation site was located. The SGP CF land cover is classified as grassland–rangeland (Crow and Wood 2003). More discussion on the potential effects of the roughness length and moisture availability on the parameterized fluxes is presented in section 5.

b. The six SFP schemes examined

The SFP schemes examined herein include three used commonly in version 3.0 of the WRF model and later [the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (Penn State–NCAR) Mesoscale Model (MM5; WRF surface layer 1), Eta Model (Eta; WRF surface layer 2), and the Pleim–Xiu land surface model (PX; WRF surface layer 7) schemes] (Skamarock et al. 2008), and those used in the three major US global climate models (GFDL, GISS, and NCAR) [GAMDT (2004) for GFDL, Schmidt et al. (2006) for GISS, and Collins et al. (2004) for CAM]. All six schemes are based on the MOST theoretical framework, but with differences in their specific treatments of stability functions, etc. For convenience, the main features of all six of the SFP schemes evaluated are summarized in Table 1, and are elaborated upon below, with an emphasis on their differences.

1) MM5 (WRF SURFACE LAYER 1) SCHEME

The MM5 scheme is designated as the surface layer 1 scheme in the WRF model, version 3.0 and later. This method uses the stability functions developed by Holtslag and De Bruin (1988) for stable conditions and by Paulson (1970) for unstable conditions. The Beljaars (1995) correction is applied to calculate the convective velocity scale, which is added to the horizontal wind speed in order to enhance the wind speed and prevent it from being zero under strong convection conditions.

According to Zhang and Anthes (1982), depending on the sign and magnitude of the bulk Richardson number R_b, the stability regime is divided into four categories, upon which either the turbulent fluxes are set to be zero or the forms of the stability profile functions $\Psi_{m,h}$ are determined. The atmospheric stability parameter z/L is obtained by solving the relation between z/L and R_b iteratively (Beljaars and Holtslag 1991).

2) ETA (WRF SURFACE LAYER 2) SCHEME

The Eta scheme is designated as the surface layer 2 scheme in WRF model, version 3.0 and later. It is based on Janjić (1996). In this scheme, the effects of the viscous sublayer are taken into account by introducing the roughness length for temperature and humidity (Zilitinkevich 1995), which is different from that for momentum. The surface fluxes are calculated by an iterative method. As in the MM5 scheme, the Beljaars (1995) correction is also applied in this scheme, as are the stability functions employ by Holtslag and De Bruin (1988) and by Paulson (1970) for stable and unstable conditions, respectively.

3) PX (WRF SURFACE LAYER 7) SCHEME

The PX scheme is designated as the surface layer 7 scheme in WRF model, version 3.0 and later. It was originally developed by Pleim (2006) and accounts for the difference in the sink–source heights of heat and momentum with the parameterization of a viscous sublayer in the form of a quasi-laminar boundary layer resistance. The stability functions are analytically estimated from large-scale state variables and are as follow.

1) For stable conditions,

   (I) When $0 < z/L < 1$, $z/L = \ln(z/z_0)[R_b/(1 - R_b/R_{\text{crit}})]$, where $R_{\text{crit}} = 0.25$ is the critical Richardson number, and $\Psi_{m,h} = -\beta_{m,h}(z/L)$, where $\beta_m = \beta_h = 1/R_{\text{crit}}$.

   (II) When $z/L > 1$, $z/L = \ln(z/z_0)[R_b/(1 - R_{\text{cut}}/R_{\text{crit}})]$, where $R_{\text{cut}} = [\ln(z/z_0) + (1/R_{\text{crit}})]^{-1}$, and $\Psi_{m,h} = 1 - \beta_{m,h} - (z/L)$.

2) For unstable conditions,

   $\Psi_{m,h} = a_{m,h}\ln\{1 - b_{m,h}[\ln(z/z_0)]^{1/2}R_b\}$, where $a_{m,h} = c_{m,h} + d_{m,h}\ln[\ln(z/z_0)]$. Here, these empirically determined constants are $b_m = 13.0$, $b_h = 15.7$, $c_m = 0.031$, $c_h = 0.04$, $d_m = 0.276$, and $d_h = 0.355$.

4) CAM SCHEME

The CAM scheme is employed in the NCAR Community Atmosphere Model (Collins et al. 2004). In the scheme, the stability functions are based on those used by Holtslag and De Bruin (1988) for stable conditions and by Paulson (1970) for unstable conditions. However, the stability functions based on Kader and Yaglom (1990) are employed for very unstable conditions and those based on Holtslag et al. (1990) are employed for very stable conditions, when atmospheric stratification and thermal effects are very strong.

The roughness lengths for momentum and heat (moisture) are estimated according to Zilitinkevich (1970) and Zeng and Dickinson (1998).The stability parameter z/L is restricted to $-100 \le (z/L) \le 2$. The scalar wind speed is defined as

$$|U|^2 = U^2 + V^2 + U_c^2,$$

$$U_c = \begin{cases} 0.1\ \text{m s}^{-1}, & \text{if} \quad \dfrac{z}{L} \ge 0\ \text{(stable)} \\ \beta w_* = \beta\left(z_i \dfrac{g}{\overline{T}}\overline{w'\theta_v'}\right)^{1/3}, & \text{if} \quad \dfrac{z}{L} < 0\ \text{(unstable)} \end{cases}.$$

Here, w_* is the convective velocity scale and $\beta = 1$; in addition, z_i is the convective boundary layer height, of which the value is taken as 1000 m.

5) GISS SCHEME

The GISS scheme is employed in the GISS model (Schmidt et al. 2006). It uses the stability functions of Holtslag and De Bruin (1988) and Paulson (1970) but with different empirical constants in the functions herein. The transfer coefficients for heat, moisture, and momentum employ the forms used by Hartke and Rind (1997). The roughness length for temperature over land is calculated according to Brutsaert (1982).

6) GFDL SCHEME

The GFDL scheme is employed in the GFDL model (GAMDT 2004). It is virtually the same as the GISS scheme, except that different empirical constants are used in the stability functions of Holtslag and De Bruin (1988) and Paulson (1970).

4. Results and analysis

a. Comparison of EBBR and EC measurements

As mentioned above, for sensible and latent heat fluxes, there are two independent sets of measurements from the EC and EBBR systems at the SGP site. To evaluate the SFP schemes within the context of observational uncertainty, we first compare the EBBR and EC observations. Figure 1 shows the sensible and latent heat fluxes from the EBBR and EC observations in the form of occurring probability density. Generally, the

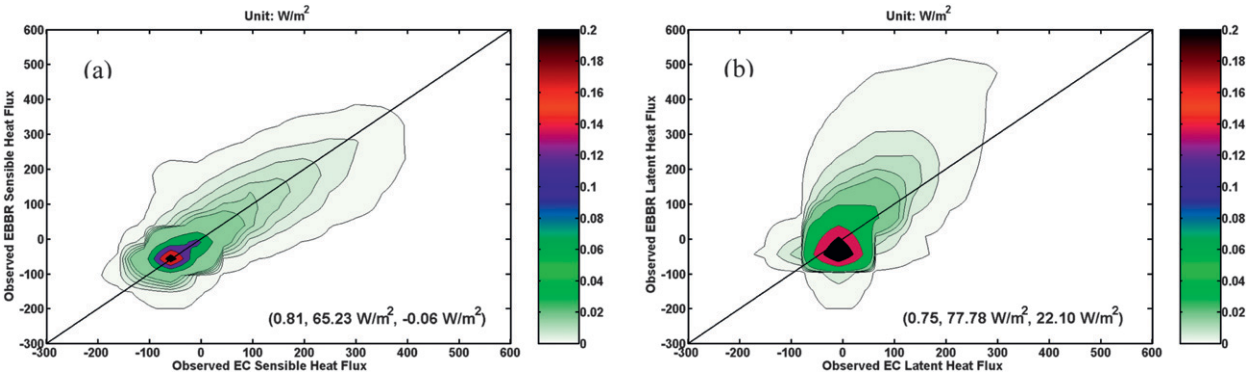

FIG. 1. Comparison of the surface (a) sensible heat and (b) latent heat fluxes between the EBBR and EC observations in the form of occurring probability density. The numbers in parentheses in each panel are the correlation coefficient, the root-mean-square error, and the bias (EBBR − EC) between the two sets of observations, respectively.

two sets of observations are in good agreement with each other, with the correlation coefficients of 0.81 and 0.75, the root-mean-square errors of 65.23 and 77.78 W m^{-2}, and the biases of −0.06 and 22.10 W m^{-2}, respectively, for the sensible and latent heat fluxes. Similar findings were reported in previous studies (Brotzge and Crawford 2003; Cook et al. 2006). The EC system appears to underestimate the latent heat flux compared to the EBBR system (Fig. 1b) in most cases, which is consistent with the finding of Brotzge and Crawford (2003). There are several possible reasons for the difference between the EC and EBBR results. One reason may lie in their different locations. The EC system is near paved surfaces rather than being field-centric at the central facility, whereas the EBBR system is closer to the wheat field. The sensors of the two systems are also located at different heights, which induce differences in the fetch and/or flux footprints ''seen'' by the instruments and create differences in the measurements.

According to Brotzge and Crawford (2003) and Cook et al. (2006), another possible reason for the differences in the fluxes between the two systems is the theoretical assumption underlying the EBBR system that the eddy diffusivities of heat and water vapor are equal. Some studies have demonstrated that the two diffusivities are not equal under stable and neutral conditions. The EBBR system may also yield biased turbulent energy fluxes toward higher latent heat fluxes due to the partitioning of the energy residual between the latent and sensible heat fluxes (Ingwersen et al. 2011). Furthermore, the EC system is known to suffer from closure problems such as the advection effect (Cook et al. 2006). More work is needed to determine the exact reasons for the discrepancies, but that level of examination is beyond the scope of this paper.

b. *Comparison of parameterizations and observations*

1) SURFACE TURBULENT FLUXES

Figure 2 compares the momentum flux (friction velocity) between the parameterizations and EC observations. It is evident that all six SFP schemes perform well relative to the EC observations, with correlation coefficients around 0.90, root-mean-square errors around 0.10 m s^{-1}, and biases around 0.03 m s^{-1}. The differences among the various SFP schemes are small.

Figure 3 compares the sensible heat flux between the parameterizations and EC observations. Unlike the momentum flux, all the schemes perform poorly, and underestimate the sensible heat flux compared to the EC observations when the observed EC fluxes are nonnegative. Relatively speaking, the MM5 scheme performs best with a correlation coefficient of 0.55 and a root-mean-square error of 125.91 W m^{-2}; the Eta scheme is the worst with a correlation coefficient of 0.36 and a root-mean-square error of 145.31 W m^{-2}. The MM5 scheme does not significantly underestimate the sensible heat flux when the observed fluxes are close to zero under stable stratification conditions. The biases between the parameterizations and EC observations are around −70 W m^{-2}, indicating an overall underestimation of the sensible heat flux by the parameterizations.

Similar to the sensible heat flux, all of the schemes perform poorly for the latent heat flux (Fig. 4). When the observed latent heat fluxes are around zero under stable conditions, the Eta and PX schemes underestimate the latent heat flux significantly. Among the six schemes, relatively, the CAM scheme and the Eta scheme are the best and worst performers, with their corresponding correlation coefficients of 0.46 and 0.38 and root-mean-square errors of 74.94 and 85.82 W m^{-2}, respectively.

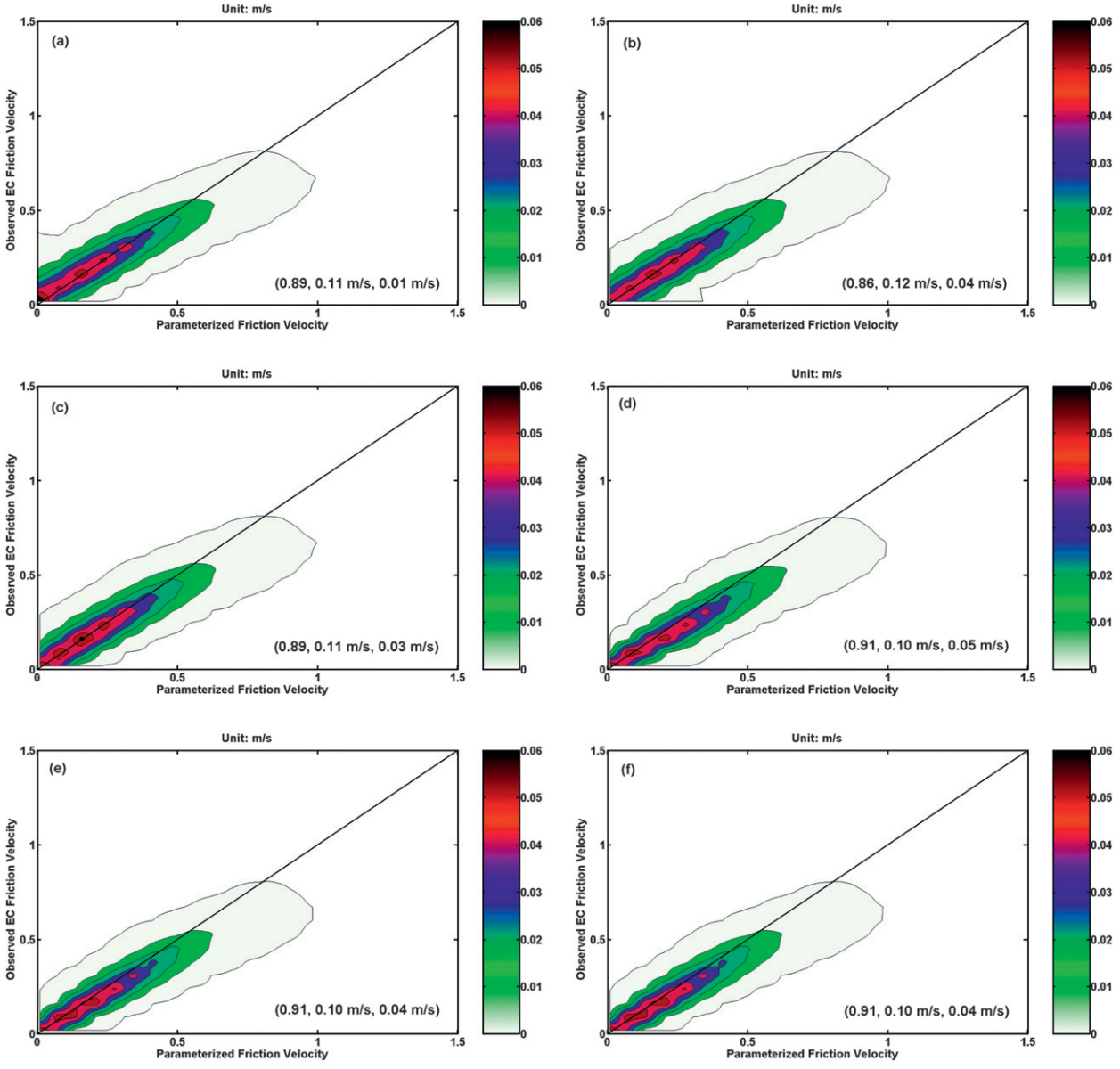

FIG. 2. Comparison of the momentum flux (friction velocity) between the EC observations and the parameterizations: (a) MM5 (WRF surface layer 1), (b) Eta (WRF surface layer 2), (c) PX (WRF surface layer 7), (d) CAM, (e) GISS, and (f) GFDL. The numbers in parentheses are as in Fig. 1.

The Eta scheme also suffers from the largest bias (-14.78 W m^{-2}).

The parameterized sensible and latent heat fluxes are also compared to the EBBR observations, and the results are similar to those in Figs. 3 and 4 and, thus, are not shown here. Briefly, the schemes underestimate the sensible heat flux when the observed fluxes are positive. In particular, the schemes underestimate the flux significantly when the observed counterparts are close to zero, except for the MM5 scheme. The MM5 and Eta schemes are, respectively, the best and worst at reproducing the sensible heat flux based on their respective correlation coefficients to the observations, which are 0.60 and 0.42. Based on the correlation coefficient and root-mean-square error, the three schemes used in the GCM models produce better estimates for the latent heat flux than do those used in the WRF model, and the Eta and PX schemes underestimate the latent heat flux significantly when the observed values are approximately equal to zero. The Eta scheme underperforms compared to the other schemes based on its correlation coefficient to the observations, which is 0.52.

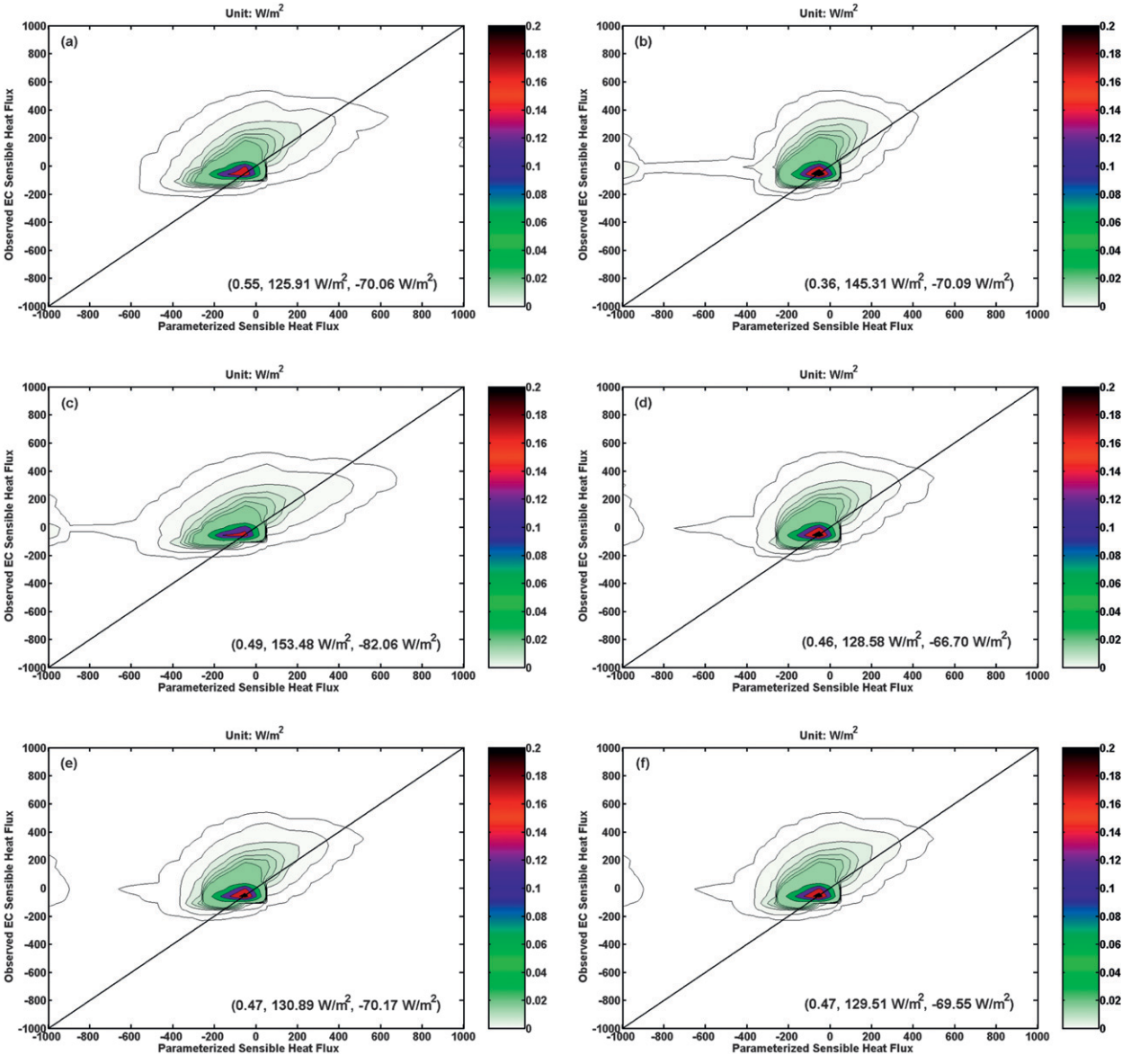

FIG. 3. As in Fig. 2, but sensible heat flux.

2) BOWEN RATIO AND EVAPORATIVE FRACTION

It has been recognized that many processes are determined by the partition between sensible and latent heat fluxes. The two relative measures commonly used to gauge this partition are the Bowen ratio (Bowen 1926) and evaporative fraction (Betts et al. 1997). The Bowen ratio is defined as the ratio of the sensible heat flux to the latent heat flux; the evaporative fraction is defined as the ratio of the latent heat flux to the sum of the sensible and latent heat fluxes. Obviously, the Bowen ratio and evaporative fraction are inversely related to each other, but are preferred by different researchers in different communities. For example, Lu and Cai (2009) showed that the fractional change of the Bowen ratio with global warming approximately follows the rate expected from the Clausius–Clapeyron equation, and is closely related to the debate on the global hydrological response to global warming. The two ratios are also essential in assessing evapotranspiration.

The above analysis shows that the parameterized sensible and latent heat fluxes exhibit different biases compared to the EC observations. In view of the unique importance of the Bowen ratio and evaporation fraction, the six SFP schemes are further evaluated against the observations in terms of the Bowen ratio and evaporative

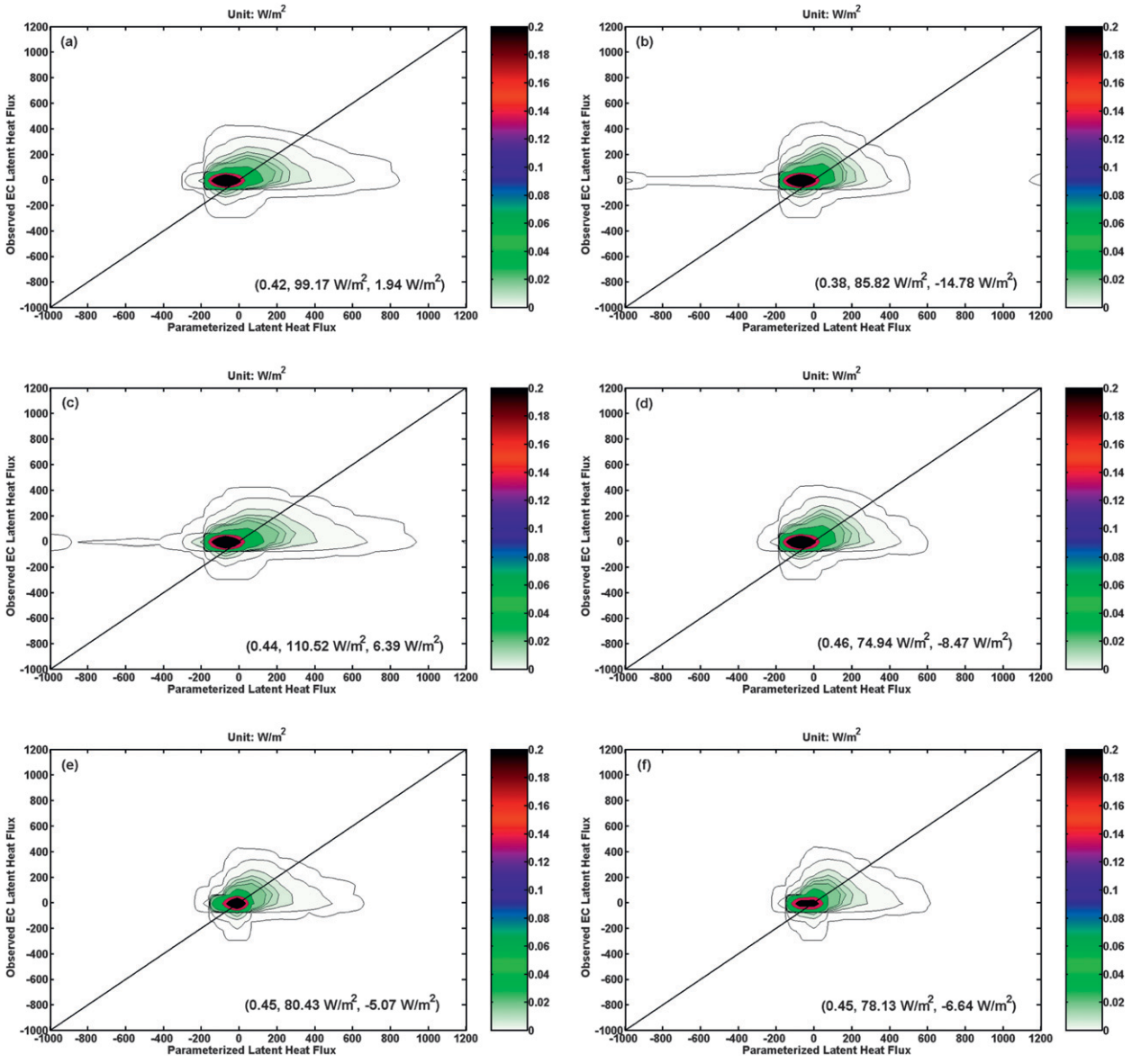

FIG. 4. As in Fig. 2, but for latent heat.

fraction. Figure 5 compares the evaporative fraction between the six SFP schemes and the EC observations. The data with $|\mathrm{LHF} + \mathrm{SHF}| \geq 10$ W m^{-2} are selected in the analysis to avoid unreasonably large values of the evaporative fraction. It is evident that all the schemes represent the evaporative ratio even more poorly than the corresponding sensible or latent heat flux. For the three schemes used in the WRF model, the correlation coefficients are as low as 0.11, the root-mean-square errors as high as 1.10 and the biases are at 0.07, while for the three schemes used in the GCM models, the corresponding quantities are respectively 0.09, 1.17, and 0.04. When the cutoff threshold increases from 10 to 50 W m^{-2}, as shown in Fig. 6, the agreement between the parameterizations and observations is improved somewhat. The correlation coefficients between the observations and the three schemes used in the WRF model increase to 0.22 and the root-mean-square errors are reduced to 0.57, while for the three schemes used in the GCM models, the two corresponding quantities become 0.17 and 0.60, respectively.

Figure 7 compares the Bowen ratio between the parameterizations and EC observations. The data with $|\mathrm{LHF}| \geq 10$ W m^{-2} are selected in the analysis to avoid having the Bowen ratio becomes unreasonably too large. The results are similar to those of the evaporative

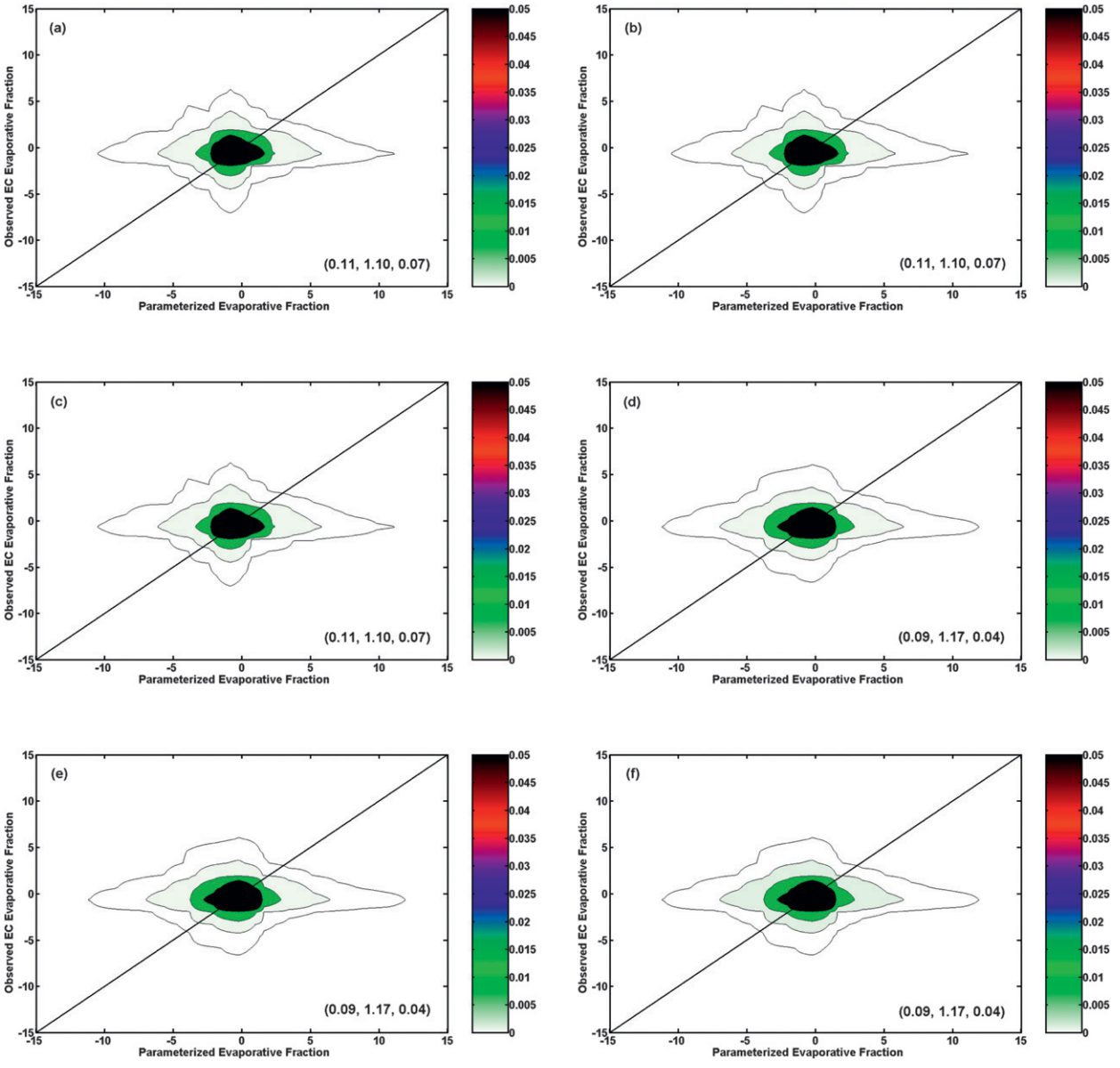

FIG. 5. As in Fig. 2, but for the evaporative fraction. The threshold value is 10 W m^{-2}, i.e., the data with |LHF + SHF| ≥ 10 W m^{-2} are selected in the analysis.

fraction. For the three schemes used in the WRF model, the correlation coefficients are 0.10, the root-mean-square errors are around 4.9, and the biases around −1.3. While for the three schemes used in the GCM models, the three corresponding quantities are 0.09, around 5.1, and around −1.2, respectively. Similar to the evaporative fraction, when the cutoff threshold increases from 10 to 50 W m^{-2}, as shown in Fig. 8, the comparisons for the Bowen ratio are improved. The correlation coefficients between the observations and the three schemes used in the WRF model increase to around 0.2 and the root-mean-square errors reduce to around 2.1, while for the three schemes used in the GCM models, the two corresponding quantities are around 0.2 and 2.2, respectively. The degradation of the parameterized evaporative fraction and the Bowen ratio reveals the magnification of the errors in the parameterized sensible and latent heat fluxes when converted into their respective ratios.

3) ANALYSIS OF TAYLOR DIAGRAM

The above analysis evaluates the performance of the six SFP schemes using the joint probability density function and in terms of the correlation coefficient,

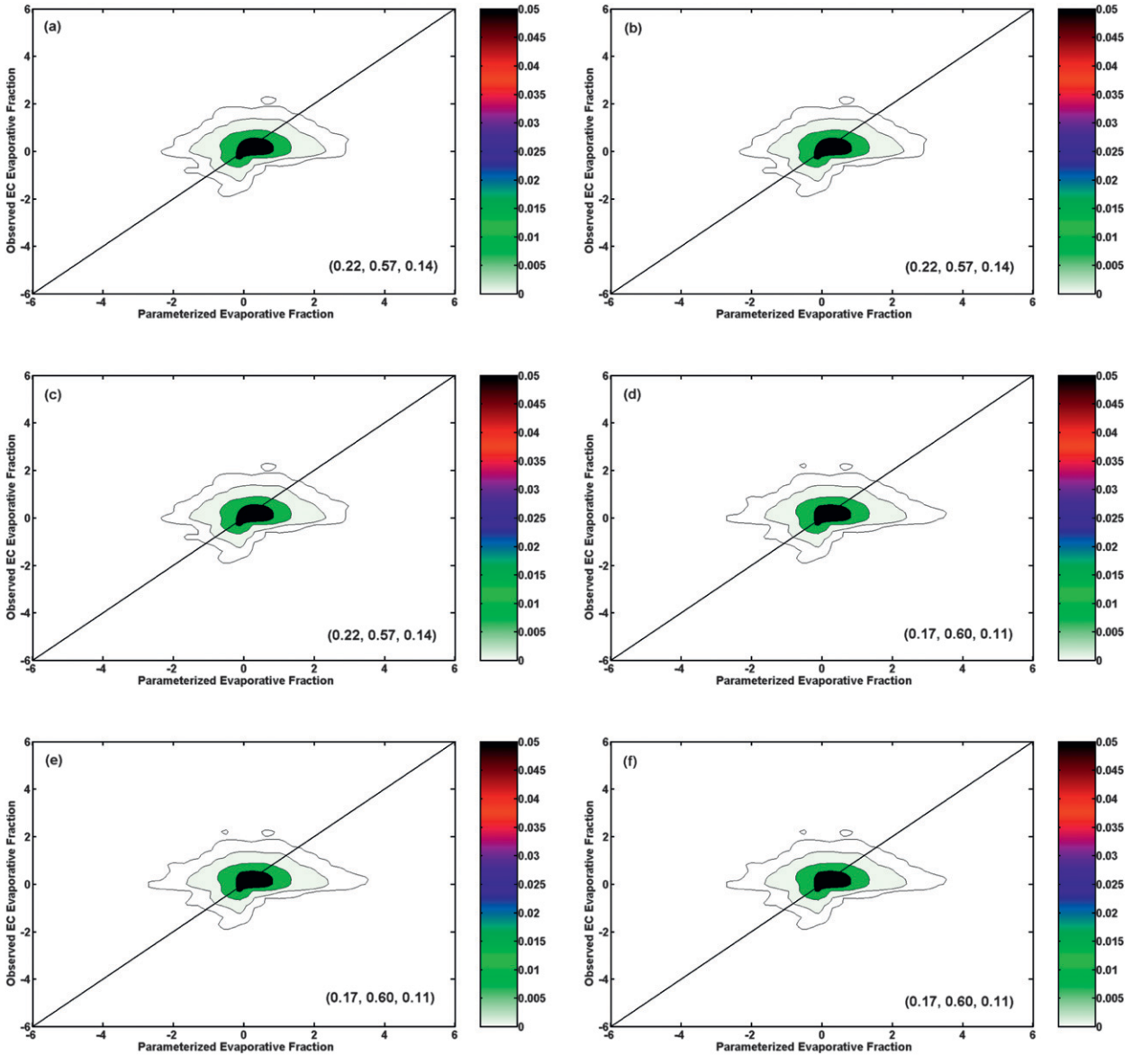

FIG. 6. As in Fig. 5, but here the threshold value is 50 W m^{-2}.

root-mean-square error, and bias. A more quantitative and complete picture of how well the parameterizations agree with the observations can be obtained by using the Taylor diagram (Taylor 2001). The Taylor diagram compares three quantities—standard deviation, correlation coefficient, and centered root-mean-square difference—in a two-dimensional plot. The angle coordinate of the Taylor diagram gives the correlation between parameterizations and observations; the radial coordinate compares the parameterized and observed amplitudes of the variations as measured by the standard deviation, and the distance between each parameterization point and the observed point gives the centered root-mean-square model error.

Figures 9a–d shows the Taylor diagrams of the momentum flux (friction velocity), sensible and latent heat fluxes, and Bowen ratio for comparisons between the parameterizations and EC observations, respectively. The EBBR observation is treated as a ''parameterization'' in the diagrams, since the EBBR flux measurement is based on the same assumption (MOST) that is employed by the six parameterizations. As shown in Fig. 9a, all the schemes lie near the point marked ''observed,'' which suggests that all six schemes parameterize the momentum flux well compared to the EC

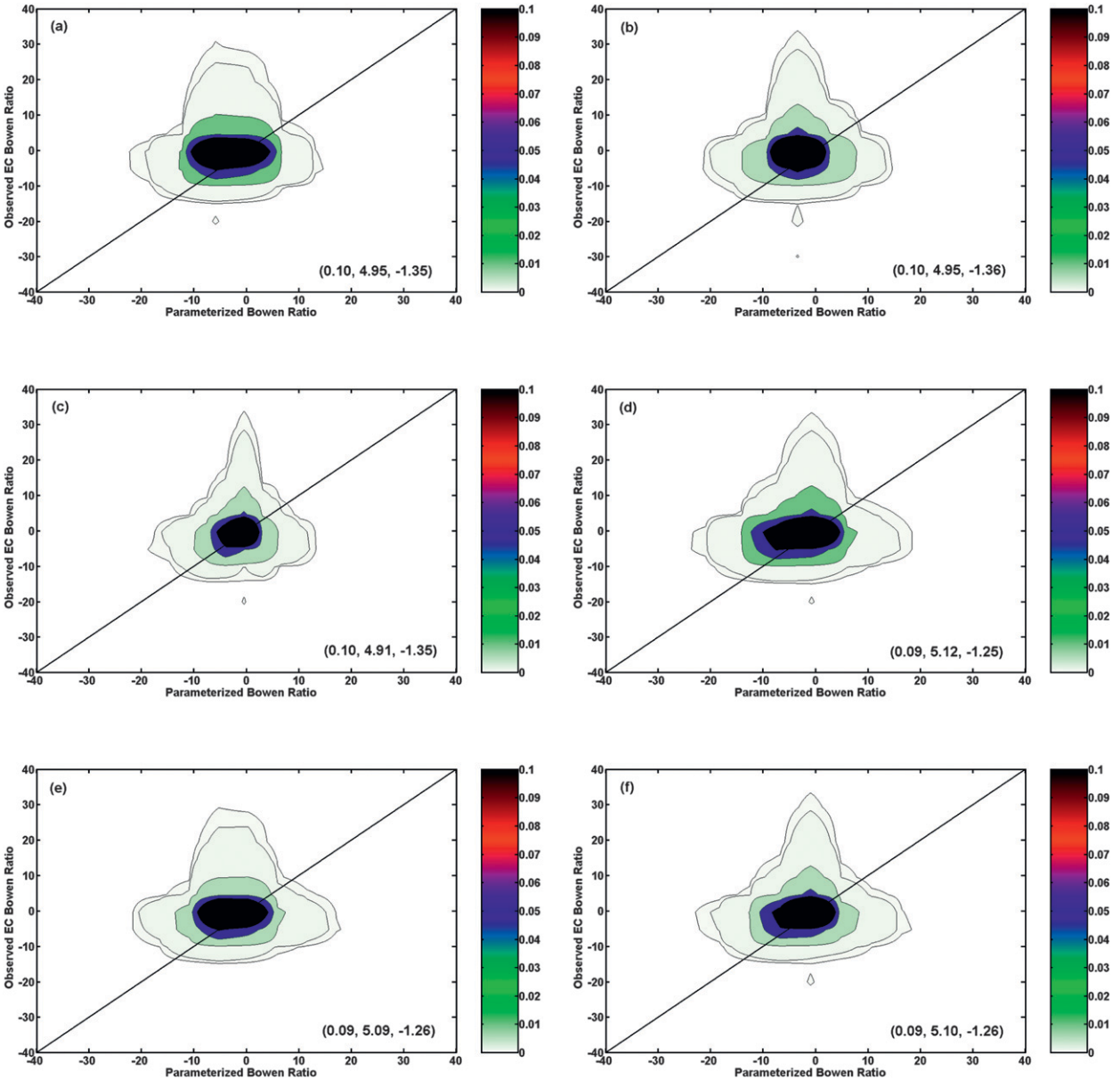

FIG. 7. Comparison of the Bowen ratio between the EC observations and the parameterizations: (a) MM5 (WRF surface layer 1), (b) Eta (WRF surface layer 2), (c) PX (WRF surface layer 7), (d) CAM, (e) GISS, and (f) GFDL. The numbers in parentheses in each panel are the correlation coefficient, the root-mean-square error, and the bias, respectively. The threshold value is 10 W m^{-2}, i.e., the data with $|\text{LHF}| \geq 10$ W m^{-2} are selected in the analysis.

observations. Moreover, the three schemes used in the GCMs have the best overall performance and their Taylor points almost overlap with one another.

For sensible heat fluxes, Fig. 9b shows that the MM5 scheme has the shortest distance to the EC observations among all the schemes, which suggests that it has the least-centered RMS error compared to the observations. Moreover, its standard deviation is closest to that of the observed, indicating the variations of its parameterization are of the correct amplitude, and it also has the largest correlation coefficient among all the schemes. Thus, the MM5 scheme is the best parameterization of the sensible heat flux compared to the EC observations. The Eta scheme is the worst, with the highest centered RMS and the smallest correlation coefficient. For latent heat fluxes (Fig. 9c), the three schemes used in the GCMs have the largest correlation coefficients, the closest standard deviations, and the smallest centered RMS errors against the EC observations, which suggests they are relatively better at parameterizing the latent

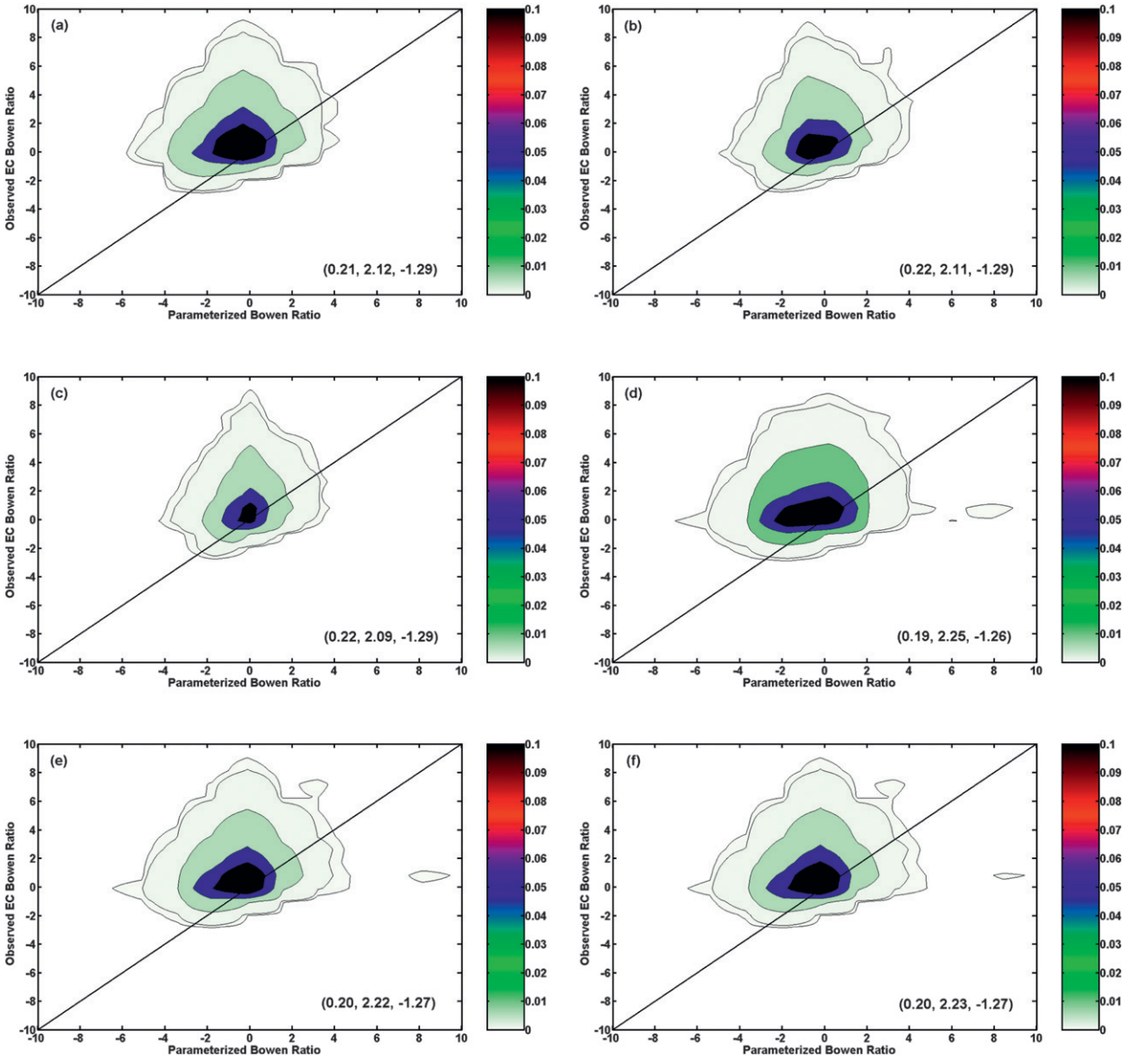

FIG. 8. As in Fig. 7, but here the threshold value is 50 W m^{-2}.

heat flux. The three schemes used in the WRF model are comparatively worse. In Fig. 9d, the points of the six schemes are almost overlapped with one another but far away from the EC observations, all of them having poor correlations and levels of agreement with the EC observations and underestimating the variability of the EC observation. This result confirms the previous finding that all six schemes poorly quantify the Bowen ratio or evaporative fraction.

4) ANALYSIS OF RELATIVE EUCLIDEAN DISTANCE

Although the Taylor diagram allows a visual comparison of three important statistics (correlation coefficient, standard deviation, and centered root-mean-square error), it ignores the mean bias, another crucial quantity in assessing any parameterization. Further, as demonstrated above, different schemes may have different levels of performance in terms of these different statistics. Further still, different quantities (e.g., momentum flux and sensible–latent heat flux) have different units. Therefore, it is desirable to have a single metric that can both measure the overall performance of a parameterization and allow for comparison of the parameterizations for different quantities. For this purpose, here we use a new dimensionless metric, the relative Euclidean distance D (Wu et al. 2012), which is defined as

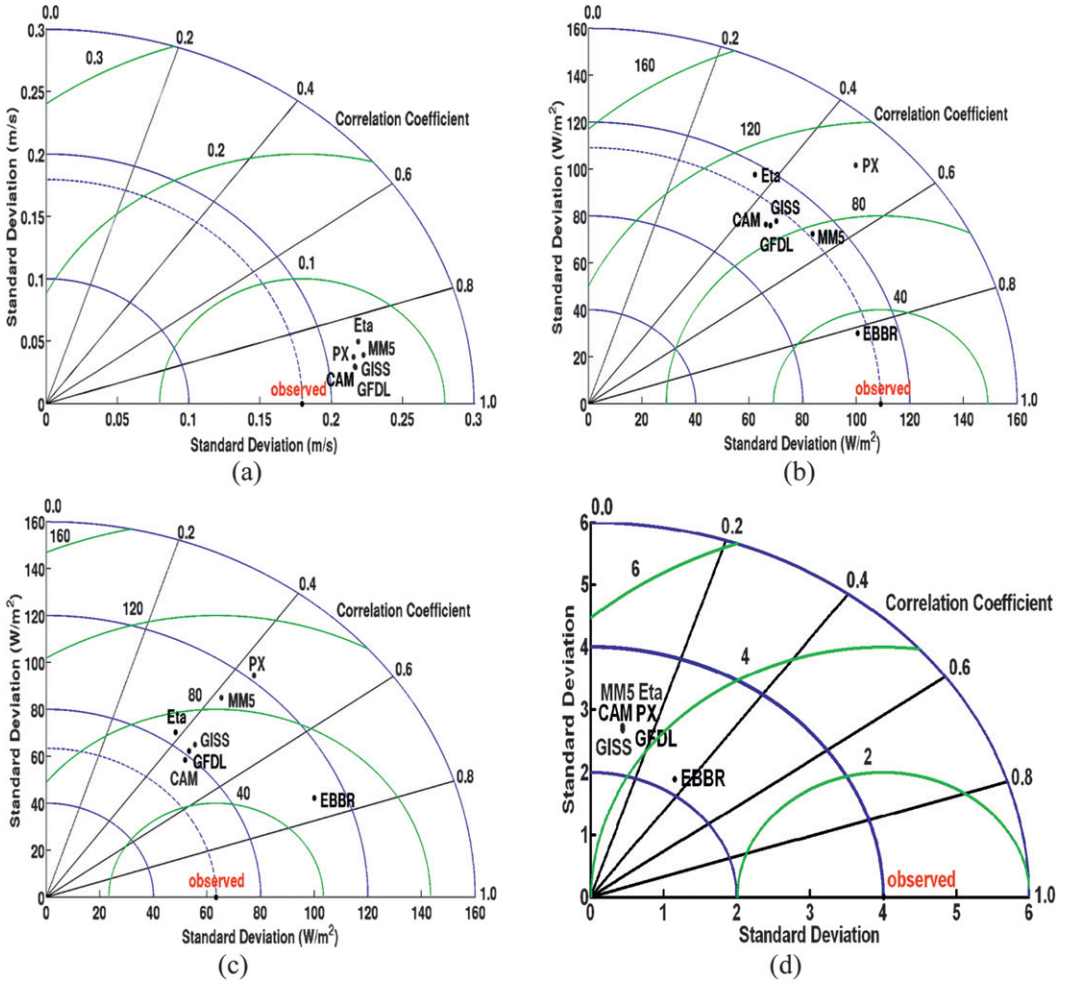

FIG. 9. Taylor diagrams of the (a) momentum flux (friction velocity), (b) sensible heat flux, (c) latent heat flux, and (d) Bowen ratio for comparisons between the parameterizations and EC observations. The EBBR observation is treated as a "parameterization" here. The blue arc represents the standard deviation of the parameterization and the observation; the green arc represents the centered root-mean-square difference between the parameterization and the observation.

$$D = \sqrt{\left(\frac{\overline{x} - \overline{y}}{\overline{y}}\right)^2 + \left(\frac{\sigma_x - \sigma_y}{\sigma_y}\right)^2 + (c_{xy} - 1)^2},$$

where x and y are, respectively, the model and observation data; $\overline{x}$, $\overline{y}$, σ_x, and σ_y are the corresponding mean values and standard deviations; and c_{xy} is the correlation coefficient between x and y. Evidently, the value of D is equal to 0 for perfect agreement and increases as the agreement degrades.

Figure 10 compares the relative Euclidean distances of the momentum, sensible and latent heat fluxes, and the Bowen ratio calculated with the cutoff threshold of 10 W m^{-2} for the six SFP schemes. The EC measurements are used as the reference and the EBBR observation is treated as a "parameterization." In terms of the relative Euclidean distance, the momentum flux is again reproduced best, and all six schemes perform very well. The sensible heat flux is reproduced least accurately. The patterns of performance among the six schemes are not that different from one another, except for the poorest performer—the PX scheme. The latent heat flux is reproduced slightly better than the sensible heat flux, and the PX scheme also performs the worst on this quantity. As a group, the three schemes used in the GCM models reproduce the latent heat flux better than do the three schemes used in the WRF model, which is in agreement with the aforementioned conclusion. The Bowen ratio is reproduced better than the sensible heat flux is, but worse than the latent heat flux is, and the relative Euclidean distances for the six schemes are almost the same, which suggests that all the schemes perform poorly on the Bowen ratio. Note that while all the data are used in the analysis of the sensible and latent

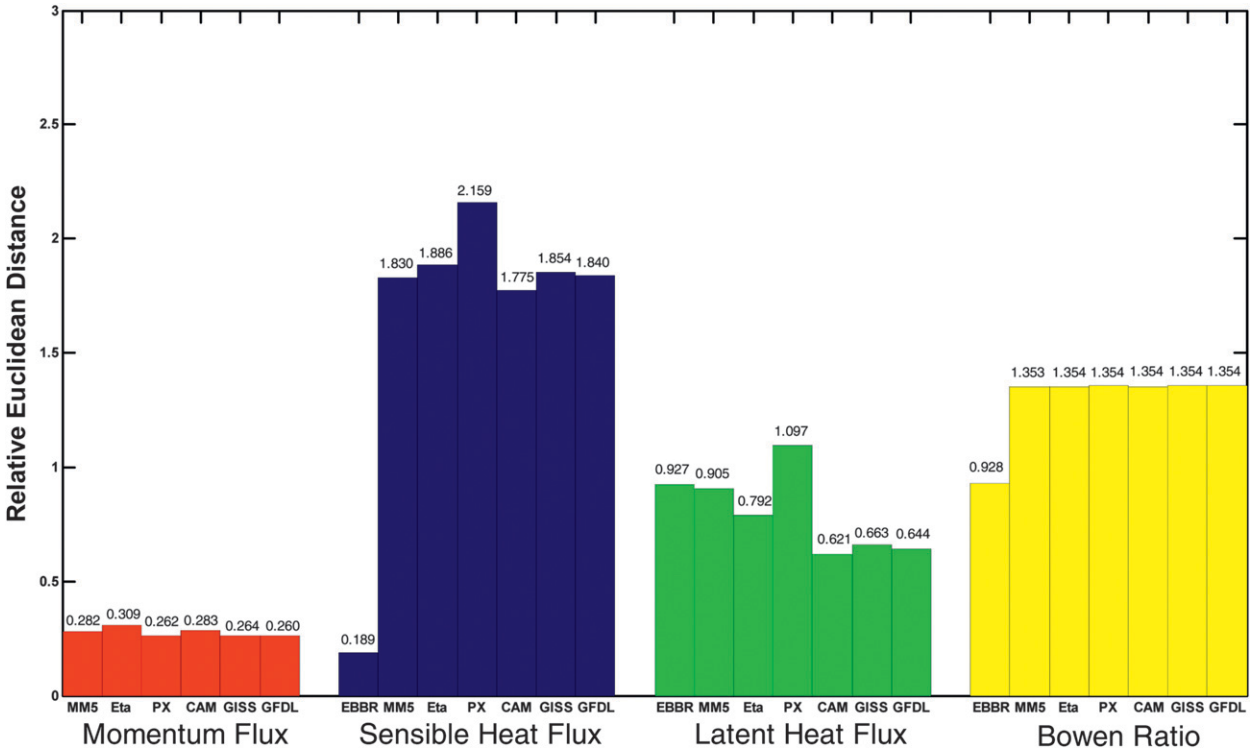

FIG. 10. Relative Euclidean distances of the momentum, sensible and latent heat fluxes, and the Bowen ratio. The EBBR observation is treated as a parameterization here.

heat fluxes, the data with the magnitude of the observed latent heat flux <10 W m^{-2}, or when the stratification is relatively stable or neutral, have been removed in the analysis of the Bowen ratio to avoid unreasonably large values of the Bowen ratio. The elimination of these data points is expected to reduce the D value of the Bowen ratio and the differences between the different schemes. It is worth emphasis that the EBBR sensible heat flux substantially outperforms all six of the SFP schemes whereas the EBBR latent heat flux is no better than the five SFP schemes, not counting the PX scheme. These results suggest that there is a need for improving the EBBR latent heat estimates and more caution must be exercised when using the EBBR latent heat flux for parameterization evaluation. For convenience, Table 2 summarizes the values used to generate Fig. 10.

5) DIURNAL VARIATION

Figure 11 further compares the diurnal variations of the momentum (panel a), sensible heat flux (panel c), and latent heat flux (panel e), respectively, in order to examine the temporal dependence of the parameterizations' performance. Also shown in Figs. 11b, 11d, and 11f are the diurnal variations of the corresponding standard deviations, respectively. Figure 11a shows that despite the excellent statistical agreement between the parameterized and EC momentum fluxes (Fig. 2), there are notable differences in their diurnal variations. Except for the PX scheme, all of the other parameterizations capture the diurnal variations to different degrees. All but the Eta and PX schemes overestimate the EC-observed momentum flux during the late morning and afternoon, but are relatively close to each other during the other times. The Eta and PX schemes always overestimate the observations; the PX scheme does not even reproduce the diurnal cycle of the observations. The Eta and PX schemes also produce the largest standard deviations during the night and early morning, as shown in Fig. 11b.

Figure 11c shows that all the parameterizations qualitatively capture, but quantitatively magnify, the diurnal cycle of the EBBR- and EC-observed sensible heat fluxes. The magnification of the diurnal cycle differs among the parameterizations, with the PX scheme being the worst. Similar patterns of behavior hold for the diurnal variation of the corresponding standard deviations. Figure 11e indicates that all the schemes are able to qualitatively capture the observed diurnal variation of the latent heat flux. The PX scheme significantly overestimates the latent heat flux in terms of the diurnal variation of the mean and standard deviation whereas the Eta scheme underestimates the latent heat flux during the night and early morning. It is noteworthy that the

TABLE 2. Summary of relative Euclidean distance (D), mean value (M), standard deviation (S), and correlation coefficient (C).

Observations/ schemes		Momentum flux	Sensible heat flux	Latent heat flux	Bowen ratio
EC	D	0	0	0	0
	M	0.310	39.527	41.541	1.759
	S	0.180	109.089	63.364	4.011
	C	1	1	1	1
EBBR	D	N/A	0.189	0.927	0.928
	M	N/A	39.462	63.641	0.900
	S	N/A	104.923	108.673	2.213
	C	N/A	0.815	0.745	0.351
MM5	D	0.282	1.830	0.905	1.353
	M	0.321	−30.538	43.478	0.062
	S	0.226	110.582	107.201	2.768
	C	0.890	0.547	0.418	0.103
Eta	D	0.309	1.886	0.792	1.354
	M	0.346	−30.560	26.764	0.060
	S	0.224	115.812	85.264	2.766
	C	0.858	0.361	0.383	0.103
PX	D	0.262	2.159	1.097	1.354
	M	0.339	−42.533	47.928	0.065
	S	0.219	142.438	122.198	2.701
	C	0.890	0.495	0.437	0.106
CAM	D	0.283	1.775	0.621	1.354
	M	0.360	−27.169	33.074	0.061
	S	0.219	101.094	78.114	2.756
	C	0.915	0.455	0.462	0.104
GISS	D	0.264	1.854	0.663	1.354
	M	0.348	−30.638	36.474	0.063
	S	0.219	104.774	85.443	2.724
	C	0.913	0.467	0.450	0.105
GFDL	D	0.260	1.840	0.644	1.354
	M	0.346	−30.027	34.898	0.063
	S	0.219	101.838	81.986	2.729
	C	0.913	0.465	0.450	0.105

interscheme differences are much larger than the two observational sets, suggesting that either EBBR or EC observations can be used to assess parameterized diurnal cycles of sensible and latent heat fluxes.

6) SEASONAL VARIATION

Figure 12 compares the seasonal variations of the momentum (panel a), sensible heat flux (panel c), and latent heat flux (panel e), as well as the seasonal variations of the corresponding standard deviations (panels b, d, and f), respectively. Figures 12a and 12b show that the seasonal variations of the parameterized and observed momentum fluxes are not that evident, while the PX and Eta schemes significantly overestimate the EC observations compared to the other schemes, and the PX and Eta schemes also have the largest standard deviations. For the sensible heat flux, Figs. 12c and 12d indicate all the schemes capture the seasonal variations of the EC and EBBR observations well in terms of both monthly mean and standard deviation, but the schemes significantly overestimate the EC and EBBR observations in the monthly mean. In Figs. 12e and 12f, the schemes also capture the seasonal variations of the EC- and EBBR-observed latent heat fluxes well, while the PX and Eta schemes overestimate the EC and EBBR observations and produce the largest standard deviations. It is noteworthy that the EC and EBBR observations reach their maxima around June and July whereas the parameterized fluxes peak in August. The lag of the parameterized latent heat fluxes is probably due to the fact that the saturation surface specific humidity at the surface skin temperature, not the actual surface specific humidity, is used in the parameterizations. More study is needed to improve the latent heat flux parameterization.

c. *Effect of atmospheric stability*

The proceeding results indicate that the discrepancies between the parameterizations and observations are related to the values of the observed fluxes and the specific times of the observations. It is well known that the sign and magnitude of surface turbulent fluxes are closely associated with the atmospheric stability, which also has strong diurnal and seasonal variations (Stull 1988). Atmospheric stability conditions play a major role in the tendency for energy and materials to move vertically through the surface layer to the free atmosphere. In an unstable atmosphere, vertical motion and turbulent fluxes are enhanced whereas vertical motion and turbulent fluxes are more likely to be suppressed in a stable atmosphere. The stability functions in the transfer coefficients of all the six schemes are derived from the MOST method, and the differences mainly lie in their specification of the stability function and the various empirical parameters embedded in the functions. As the MOST approach estimates turbulent exchanges for scalar and momentum fluxes and describes the relationship between the turbulent statistical quantities and the mean meteorological quantities in the surface layer, a MOST-based parameterization is expected to be closely related to the atmospheric stability.

To investigate the effect of the atmospheric stability on the performances of the SFP schemes, the comparisons between the parameterizations and observations are shown in terms of atmospheric stability classifications. In this study, the atmospheric stability is divided into 13 classifications according to the values of the atmospheric stability parameter z/L); these classifications span from less than −5 to larger than 5. The classifications are divided equally between −5 and 5, except that the values between −0.01 and 0.01 are set as a classification that represents the near-neutral conditions. Figure 13 compares the mean relative errors of the parameterizations as a function of the atmospheric

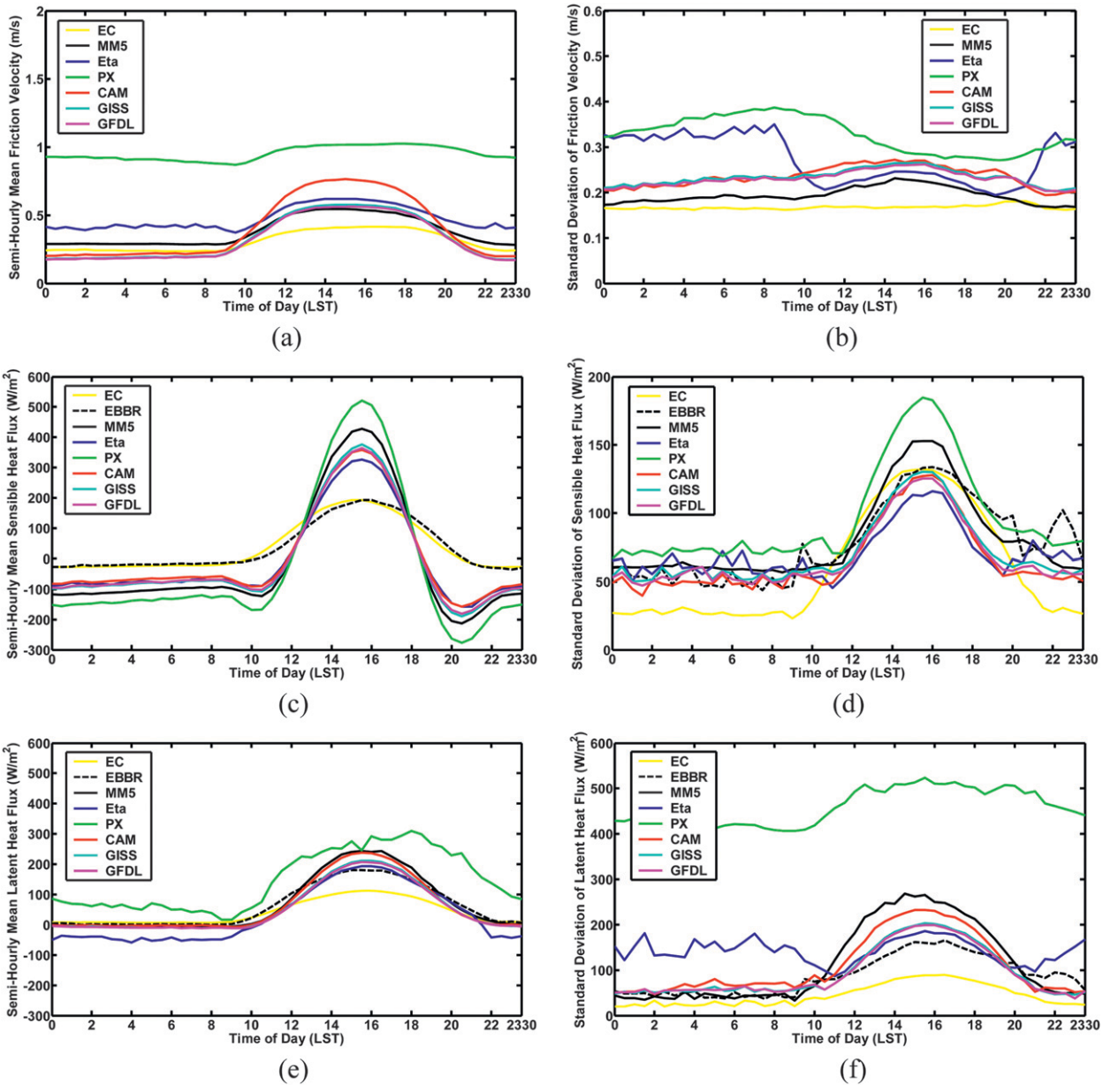

FIG. 11. Comparison of diurnal variation of the surface turbulent fluxes between the parameterizations and EC observations: (a) semihourly mean of the momentum flux (friction velocity), (b) standard deviation of the semihourly mean of the momentum flux (friction velocity), (c) semihourly mean of the sensible heat flux, (d) standard deviation of the semihourly mean of the sensible heat flux, (e) semihourly mean of the latent heat flux, and (f) standard deviation of the semihourly mean of the latent heat flux. The EBBR observation is treated as a parameterization here.

stability. The mean relative error is calculated as the averaged value of $(x - y)/y$ within an atmospheric stability classification, where x is the parameterized value and y is the corresponding observed EC value. A negative (positive) value of the relative difference indicates an underestimation (overestimation) by the parameterization compared to the observation when the observed value is positive. The EBBR observation is treated here as a parameterization. Figure 13a shows that all of the SFP schemes somewhat overestimate the momentum flux compared to the observations, and they display the same error pattern of increasing when the atmosphere becomes more unstable or more stable from the minimum near the neutral atmosphere. Different schemes perform differently under different conditions of atmospheric stability. The MM5 scheme has the smallest errors under unstable

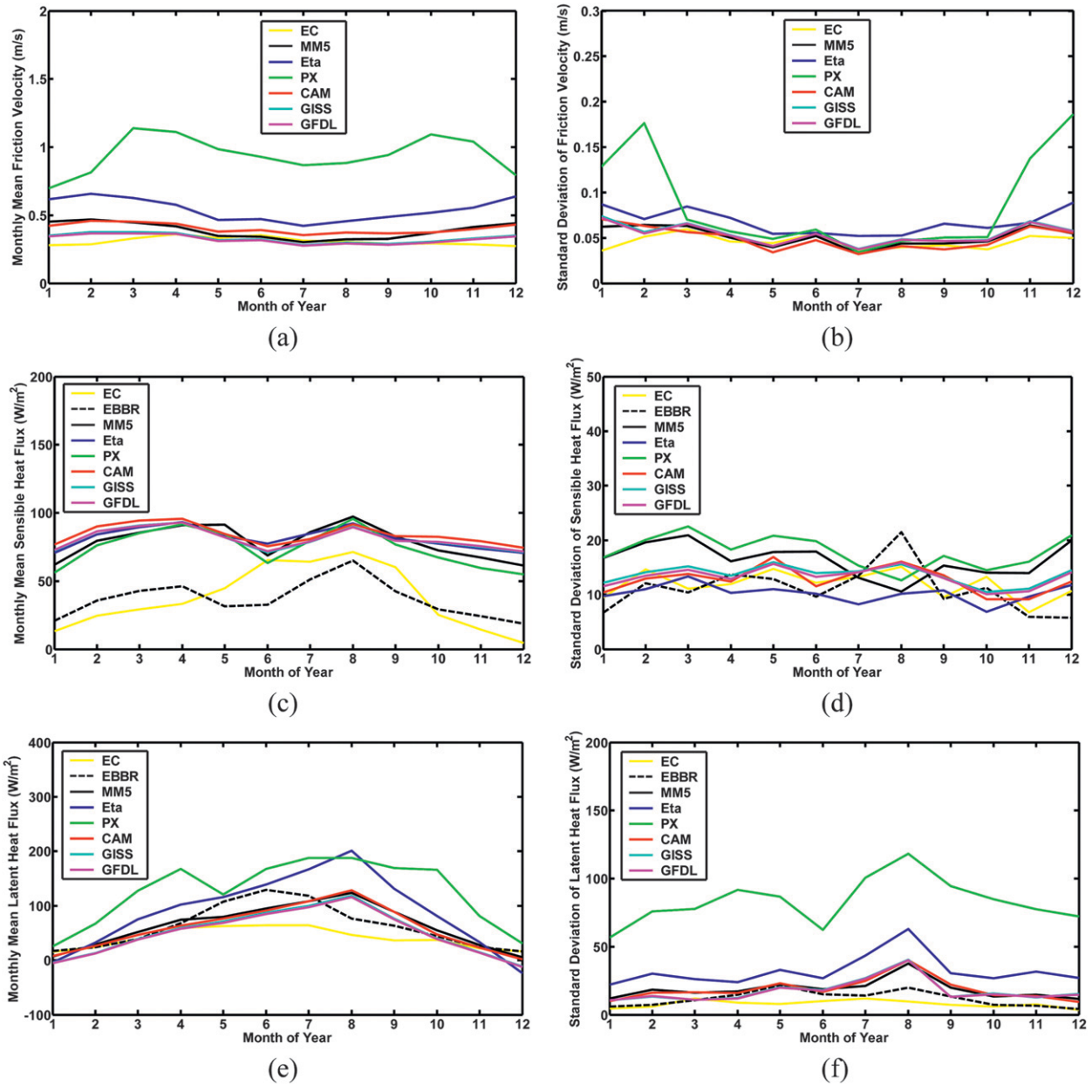

FIG. 12. Comparison of seasonal variation of the surface turbulent fluxes between the parameterizations and EC observations: (a) monthly mean of the momentum flux (friction velocity), (b) standard deviation of the monthly mean of the momentum flux (friction velocity), (c) monthly mean of the sensible heat flux, (d) standard deviation of the monthly mean of the sensible heat flux, (e) monthly mean of the latent heat flux, and (f) standard deviation of the monthly mean of the latent heat flux. The EBBR observation is treated as a parameterization here.

conditions, while the Eta scheme has the largest errors under stable conditions. But the differences between the six schemes are not large in general. Figure 13b shows that the SFP schemes underestimate the sensible heat flux under unstable conditions. Note that all the schemes also underestimate the sensible heat flux under stable conditions since the observed sensible heat fluxes are negative under those conditions. The MM5 scheme exhibits the best performance with the smallest mean relative errors under most atmospheric stability conditions, while the Eta scheme performs the worst, especially under moderate stable conditions. It is also shown that the EBBR observations underestimate the sensible heat flux compared to the EC observations under stable conditions. As shown in Fig. 13c, when the stability increases, the schemes appear to underestimate the flux. However,

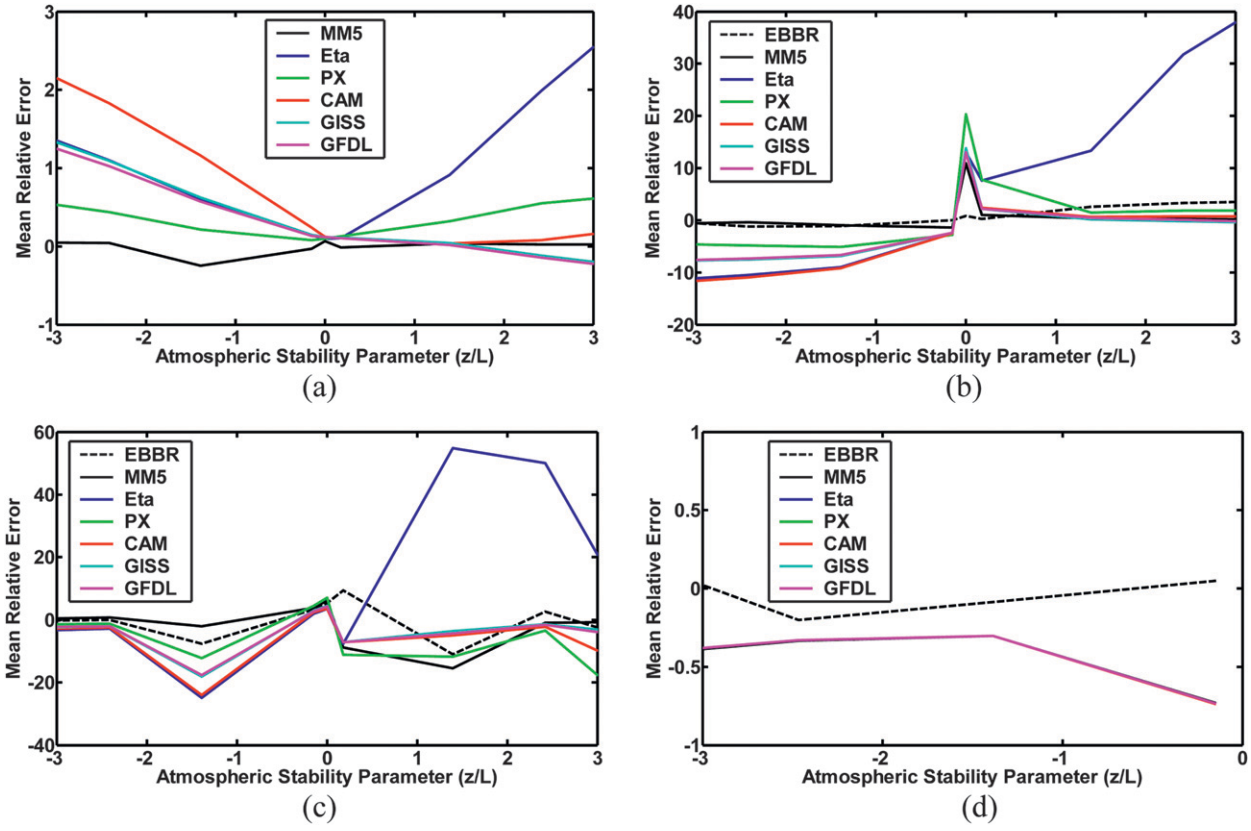

FIG. 13. Comparison of the (a) momentum flux, (b) sensible heat flux, (c) latent heat flux, and (d) Bowen ratio between the parameterizations and EC observations in the form of the mean relative error varying with the atmospheric stability. See the text for the definition of the mean relative error shown in the fractional error in each panel. The EBBR observation is treated as a parameterization here.

except for the Eta scheme, the parameterizations tend to overestimate the flux under stable conditions since the observed fluxes are negative under stable conditions. Like the sensible heat flux, the Eta scheme suffers from the largest mean relative errors under most atmospheric stability conditions, especially under moderately stable conditions. Also, the EBBR observations overestimate the latent heat flux compared to the EC observations mainly under stable conditions. Figure 13d compares the results for the Bowen ratio. Since the data with $|\mathrm{LHF}| < 50\ \mathrm{W\ m^{-2}}$ are removed in the analysis here when the stratification is relatively stable or neutral, which literally removes the data with the stability parameter larger than 0, the comparison is plotted for oderately unstable conditions only. We can see that all the SFP schemes underestimate the Bowen ratio, and they also nearly overlap with one another, which is consistent with what is seen in Figs. 9d and 10; while the EBBR observation is close to the ratio. Nevertheless, based on the results for the sensible and latent heat fluxes in Figs. 13b and 13c, the relative errors are very large when the stability parameter is larger than 0.

A collective errors analysis of the parameterizations for momentum and the sensible heat and latent heat fluxes as a function of stability reveals that the MOST-based SFP schemes work the best near neutral atmospheric conditions but degrade as the atmosphere becomes either stable or unstable. This result is consistent with the idea that the MOST approach describes the surface layer most accurately under neutral conditions. The problem with the MOST scheme beyond neutral conditions (in particular, for strong stable conditions) is well known, and many efforts have been devoted to expanding the applicability MOST method (e.g., Jiménez et al. 2012). It is worth mentioning that the majority of atmospheric stability lies between -1 and 1 (Fig. 14), suggesting that the atmosphere is dominantly under weakly unstable, near-neutral, and weakly stable conditions.

5. Other possible factors for poor parameterized sensible and latent heat fluxes

In addition to the atmospheric stability, other reasons for the poor performance of the sensible and latent heat flux parameterizations include the uncertainty–error in

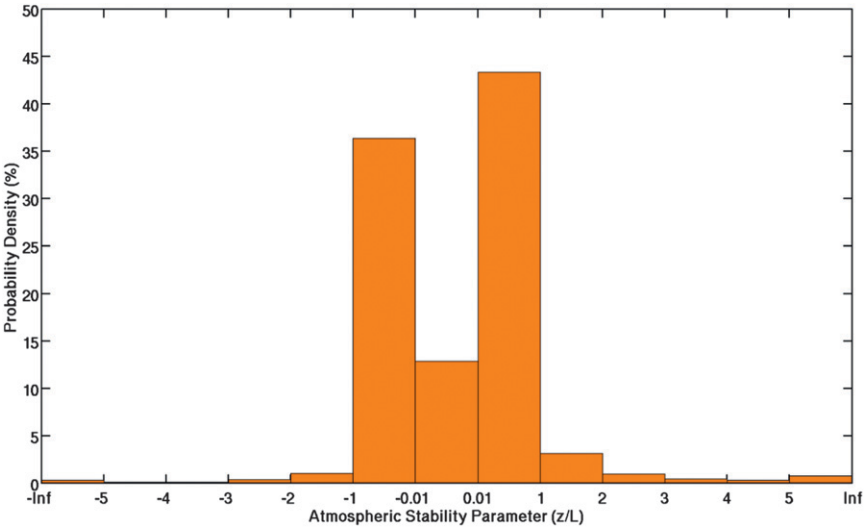

FIG. 14. Frequency distribution of the atmospheric stability parameter.

the EC and EBBR observations and the specification and accuracy of the observed surface temperature, moisture availability, and roughness length.

According to Zhan et al. (1996), errors in predicted sensible heat fluxes are predominantly sensitive to errors in air and surface temperatures. A 10% error in surface temperature (°C) can result in over 50% error in the predicted sensible heat flux. The surface temperature is also needed to determine the saturation surface specific humidity in the SFP schemes. Theoretically, the aerodynamic temperature at the thermal roughness height should be used in the calculation of sensible and latent heat fluxes. However, because aerodynamic temperature was not measured, the observed radiative surface skin temperature is used here instead to calculate the parameterized sensible and latent fluxes. The use of the skin temperature may be partly responsible for the poor performance of the parameterizations for sensible and latent heat fluxes compared to the observations [in particular, the much stronger diurnal cycle of the parameterized sensible heat fluxes discussed in section 4b(5)].

It is expected that skin temperature is closely related to the surface net radiation flux (Garratt 1992). To further uncover the possible link between using skin temperature and errors in the parameterized heat fluxes, we examine the diurnal covariations of the surface net radiation observed during the same period of time, the difference between the surface radiative and 3-m air temperatures, and the difference between the parameterized and EC-observed fluxes (Fig. 15). As expected, both the temperature and flux differences vary virtually in phase with the net radiation flux. The maximal differences appear approximately at the same time as when the net radiation is largest. However, the maximal negative differences appear around the time when the net radiation approaches zero (i.e., during the transitional period). These results further reinforce that the use of the surface radiative temperature in place of the aerodynamic temperature increases the amplitude of the parameterized surface fluxes and is likely a major reason for the substantial differences between the measured and parameterized heat fluxes. Improved understanding and quantification of the relationship between skin temperature and aerodynamic temperature is in order.

It has been known since the 1970s that the radiative surface temperature is not the same as the aerodynamic surface temperature, and many studies have been devoted to correcting for the difference between the aerodynamic and radiative temperatures (Garratt and Hicks 1973; Lhomme et al. 1994; Chehbouni et al. 1996, 1997, 2001; Sun 1999; Kustas et al. 1989, 2007; Zibognon et al. 2002; Colaizzi et al. 2004; Merlin and Chehbouni 2004). The correction methods can be divided into two general groups. One group attempts to account directly for the difference between the two temperatures. For example, Chehbouni et al. (1996, 1997) related the temperature difference to the leaf area index (LAI) such that

$$\beta = \frac{T_o - T_a}{T_r - T_a} = \frac{1}{\exp[\mathrm{EF}/(\mathrm{EF} - \mathrm{LAI})] - 1}, \qquad (12)$$

where T_o is the aerodynamic surface temperature, T_a is the air temperature at a reference height above the surface, T_r is the radiative surface temperature, and EF is an empirical factor that may depend on the vegetation type and structure and was set to a value of 1.5. Sun (1999) found by empirical analysis of data collected over grasslands that the two temperatures were linearly related to each other.

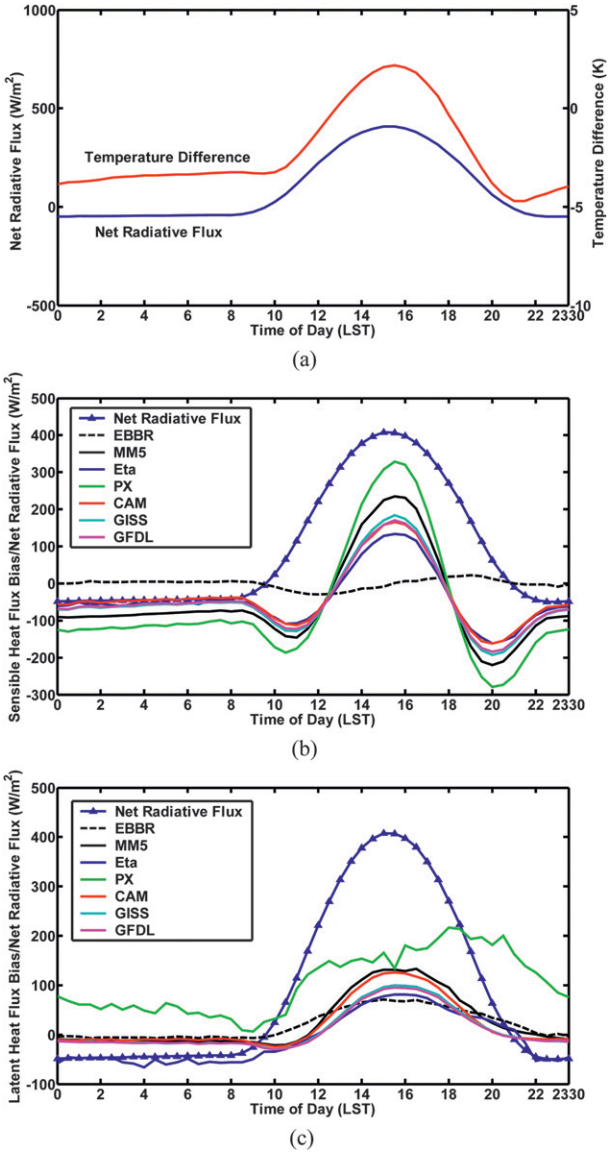

FIG. 15. Semihourly mean of the temperature and flux differences varying with the surface net radiation. (a) The temperature difference is the surface radiative temperature minus the air temperature, and (b),(c) the flux difference is the parameterized surface flux minus the EC-observed surface flux. The EBBR observation is treated as a parameterization here.

The other method is to add a correction term to the aerodynamic resistance in the formulation to compute the surface fluxes (Kustas et al. 1989, 2007; Colaizzi et al. 2004):

$$R_{\mathrm{EX}} = \frac{\ln(z_{0\mathrm{m}}/z_{0\mathrm{h}})\{\ln[(z-d)/z_{0\mathrm{m}}] - \psi_m\}}{k^2 u}, \quad (13)$$

where u is the wind speed at the height z above the surface; k (=0.4) is the von Kármán constant; d is the displacement height; and $z_{0\mathrm{h}}$ are the roughness lengths for momentum and heat, respectively; and ψ_m is the stability function for momentum. Specification of $z_{0\mathrm{m}}$ and $z_{0\mathrm{h}}$ further requires detailed information on surface properties such as vegetation structure and land cover type.

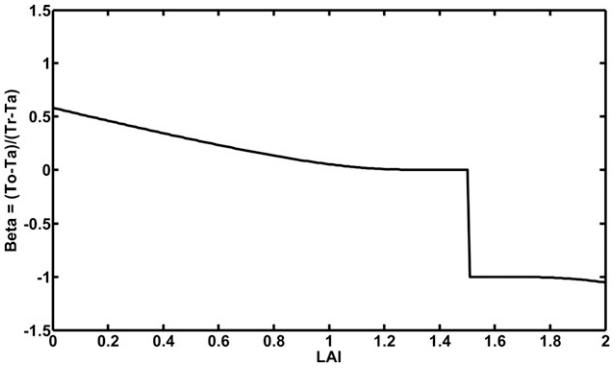

FIG. 16. Dependence of the temperature correction factor β on the LAI calculated from Eq. (11) with LAI varying from 0 to 2.

Lacking data on detailed surface properties such as vegetation structure, we are not able to implement Eq. (13) or other correction schemes. Instead, we use Eq. (12) to estimate the possible range of errors in the surface fluxes incurred by the replacement of T_o with T_r. Sun (1999) recommended the first method because of the complexity involved in determining the thermal roughness length. According to Hollinger and Daughtry (1999) and Santanello et al. (2007), the typical value of the leaf area index in summer at SGP is 2.0, suggesting a likely range of LAIs between 0 and 2 (from no vegetation to summer mature growth) at SGP. Without LAI measurements at SGP, we use Eq. (12) with LAI = 0 and 2 to estimate the possible range of errors.

Figure 16 shows the relative difference for LAI ~ 0–2. Equation (12) with a leaf area index of 2.0 yields an error E of about 5%. Since the empirical factor EF is in question, the results should be treated as empirical values for the relative difference $(T_o - T_a)/(T_r - T_a)$, and negative values do not accurately mean $T_o - T_a$ is in reverse sign to $T_r - T_a$. While estimation from Eq. (12) may be taken as a first approximation for the issue, it must be used cautiously.

It is also known that the issue is highly complex, since the difference between the two temperatures depends on many factors, including atmospheric stability, solar zenith angle, surface soil moisture, vegetation status, etc. Although some simulations have indicated that solar radiation and leaf area strongly affect the magnitude of the temperature difference, the relationship between the two temperatures is nonunique and exhibits significant variability depending on local conditions (e.g., Chehbouni et al. 1996; Kustas et al. 2007). An ultimate solution

demands an improved land surface model that accurately describes–predicts the required surface properties, which is a formidable challenge and beyond the scope of this paper.

In addition to the stability and difference between the aerodynamic and radiative skin temperatures, Eq. (5) indicates that the parameterized latent heat flux is also affected by water availability. Another potential factor is the momentum roughness length and its dependence on surface properties such as canopy structure, though Sun (1999) reported that the moment roughness remained approximately unchanged at the diurnal scale. Although water availability is expected to depend on soil moisture and vegetation details (Ye and Jia 1991), its minimal diurnal variation is unlikely the main reason for the much stronger diurnal variation in the latent heat flux. The difference in the diurnal cycle of the momentum flux between parameterized and observed fluxes also points to another culprit. Further investigation with improved data on the surface properties is in order.

6. Conclusions

The long-term (2003–10) observations of surface momentum and sensible heat and latent heat fluxes collected with the EC and EBBR systems at the Southern Great Plains site by the DOE ARM program are used to evaluate the six surface flux parameterization schemes commonly used in the Weather Research and Forecasting (WRF) model and three U.S. global climate models (GFDL, GISS, and NCAR). The schemes are assessed in terms of their performances in quantifying the correlation coefficient, variability as measured by standard deviation, centered root-mean-square error, and mean bias using an integrative analysis of joint occurrence frequency, Taylor diagrams, and the newly introduced relative Euclidean distance. Also examined are the diurnal and seasonal characteristics of the observed and parameterized surface fluxes, and the effects of the atmospheric stability and the use of surface radiative temperature to replace the aerodynamic temperature in the parameterization schemes. The main results are summarized below.

Statistical analysis shows that among the quantities examined (momentum flux, sensible heat flux, latent heat flux, Bowen ratio, and evaporation fraction), the best parameterized is the momentum flux. All six SFP schemes perform well with parameterized momentum fluxes with only a small discrepancy between the different schemes. Nevertheless, there still are notable differences in the diurnal cycle and in the functional dependence on stability, suggesting the need for further improvement.

The sensible and latent heat fluxes observed by the EBBR and EC systems are in reasonably good agreement with each other, although the discrepancy is still noteworthy. The parameterized sensible heat and latent heat fluxes compare poorly with the corresponding EC observations and all six of the SFP schemes underestimate the sensible heat flux when the observed fluxes are positive. Relatively, the three schemes used in the GCMs produce better estimates for the latent heat flux than do those used in the WRF model. Furthermore, all the parameterization schemes tend to exaggerate the magnitude of the diurnal variation of the sensible heat flux, although they qualitatively capture the diurnal cycle. All the schemes also qualitatively reproduce the diurnal cycle of the latent heat flux. While the PX scheme overestimates the latent heat flux all time, the other schemes overestimate mostly during the late morning and afternoon, and the Eta scheme underestimates during the night and early morning.

All of the parameterization schemes capture the seasonal variations of the sensible and latent heat fluxes, but they significantly overestimate the sensible heat flux in all months. Moreover, the seasonal maximum of the parameterized latent heat fluxes is lagged for about 1 month compared to the EC and EBBR observations. The PX and Eta schemes also have the largest monthly mean values of the momentum and latent heat fluxes. The errors in the parameterized sensible and latent heat fluxes are further magnified when they are converted into their respective Bowen ratio or evaporative fraction, presenting higher accuracy requirements for the SFP schemes.

Inspection of the dependence of the SFP schemes on the atmospheric stability reveals the following points. First, the difference between the EC-observed momentum flux and the parameterized counterparts reaches its minimum near neutral conditions, and becomes increasingly larger when the atmosphere becomes more stable or more unstable. Second, compared to the EC-observed sensible heat flux, the MM5 scheme has the best performance with the least mean relative errors under most of atmospheric stability conditions, while the Eta scheme does the worst, especially under stable conditions. Third, the SFP schemes tend to overestimate the latent heat flux under both strongly stable and strongly unstable conditions. Fourth, even without considering the data under stable and neutral stratification conditions to avoid unreasonably large values of the Bowen ratio, none of the schemes are able to reproduce the observations well, and the differences between the different schemes are small. Finally, the EBBR observations underestimate the sensible heat flux but overestimate the latent heat flux under stable

conditions compared to the EC observations, which is consistent with the findings of Brotzge and Crawford (2003).

Collectively, these of the results found here suggest that further parameterization improvement requires improving the common MOST theoretical framework itself. The study also suggests that the MOST method works relatively better under convective and neutral conditions than under stable conditions, which is consistent with many previous researchers' conclusions (e.g., Derbyshire 1995; Hill 1997; Mahrt 1998, 1999; Pahlow et al. 2001; also see the special issues of *J. Atmos. Sci.*, 2003, Vol. 60, No. 20, and *Bound.-Layer Meteor.,* 2006, Vol. 118, No. 2). The SFP schemes under stable conditions warrant special attention. Further diagnostic analysis reveals that the biases of parameterized sensible and latent heat fluxes, the difference between the surface radiative temperature and the air temperature measured at 3-m height, and the net radiative flux vary virtually in phase with one another, suggesting that using the radiative skin temperature to replace the aerodynamic temperature in the evaluation is at least partly responsible for the poor performance of the parameterized sensible and latent heat fluxes. Lack of accurate moisture availability and roughness length may also contribute to the poor performance of the parameterization schemes.

Therefore, the actual differences between the parameterized and observed heat fluxes may not be as large as they appear in this study because of the use of radiative skin temperature in place of aerodynamic temperature, the specified water availability, and the roughness length used in the calculation of the parameterized heat fluxes. More adequate evaluation of the surface flux parameterizations calls for improving the consideration of all of these input factors, which demands improved land surface models that incorporate complex vegetation responses, detailed hydrology, dynamical snowpack evolution, and more, a daunting challenge in itself (van den Hurk et al. 2012). Equally challenging is improving the measurements, which would allow for observational evaluation of such enhanced land surface models. Furthermore, both the measurements of the fluxes and variables used as inputs to the parameterization schemes such as skin temperature suffer from measurement uncertainties–errors, which likely compromise the results somewhat. In-depth analysis of measurement uncertainty and error propagation in parameterized fluxes is needed. The potential mismatch between the measurements and the model scales also calls for analysis of measurement representativeness.

The relatively poor performance of the parameterizations for sensible and latent heat fluxes deserves special emphasis in view of their close coupling with the boundary layer and cloud processes (Santanello et al. 2007; Betts 2009). We plan to perform online full model evaluations to further delineate the deficiencies in the surface flux parameterizations, and the coupling between the land surface, boundary layer, and cloud processes.

Acknowledgments. This work is part of the FASTER project (www.bnl.gov/esm) supported by the DOE Earth System Modeling (ESM) program. The first author is partly funded by National Basic Research Program of China (973 Program) under Grant 2010CB428501 for the work. Discussions with Drs. Alan Betts and David Cook are appreciated.

REFERENCES

Ackerman, T. P., and G. M. Stokes, 2003: The atmospheric radiation measurement program. *Phys. Today,* **56,** 38–44, doi.org/10.1063/1.1554135.

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: On flux parameterization over land surfaces for atmospheric models. *J. Appl. Meteor.,* **30,** 327–341.

Betts, A., 2009: Land-surface–atmosphere coupling in observations and models. *J. Adv. Model. Earth Syst.,* **1,** 1–18.

——, F. Chen, K. Mitchell, and Z. Janjić, 1997: Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta Model using FIFE data. *Mon. Wea. Rev.,* **125,** 2896–2916.

Bowen, I. S., 1926: The ratio of heat losses by conduction and by evaporation from any water surface. *Phys. Rev.,* **27,** 779–787.

Brotzge, J. A., and K. C. Crawford, 2003: Examination of the surface energy budget: A comparison of eddy correlation and Bowen ratio measurement systems. *J. Hydrometeor.,* **4,** 160–178.

Brown, A. R., and Coauthors, 2002: Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quart. J. Roy. Meteor. Soc.,* **128,** 1075–1093.

Brutsaert, W. H., 1982: *Evaporation into the Atmosphere: Theory, History and Applications*. D. Reidel, 299 pp.

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.

Cassano, J. J., T. R. Parish, and J. C. King, 2001: Evaluation of turbulent surface flux parameterizations for the stable surface layer over Halley, Antarctica. *Mon. Wea. Rev.,* **129,** 26–46.

Chehbouni, A., D. Lo Seen, E. G. Njoku, and B. M. Monteny, 1996: Examination of the difference between radiative and aerodynamic surface temperatures over sparsely vegetated surfaces. *Remote Sens. Environ.,* **58,** 177–186.

——, ——, ——, J.-P. Lhomme, B. Monteny, and Y. H. Kerr, 1997: Estimation of sensible heat flux over sparsely vegetated surfaces. *J. Hydrol.,* **188–189,** 855–868.

——, Y. Nouvellon, J.-P. Lhomme, C. Watts, G. Boulet, Y. H. Kerr, M. S. Moran, and D. C. Goodrich, 2001: Estimation of surface sensible heat flux using dual angle observations of radiative surface temperature. *Agric. For. Meteor.,* **108,** 55–65.

Chen, F., Z. Janjić, and K. Mitchell, 1997: Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta Model. *Bound.-Layer Meteor.,* **85,** 391–421.

Colaizzi, P. D., S. R. Evetta, T. A. Howella, and J. A. Tolka, 2004: Comparison of aerodynamic and radiometric surface temperature using precision weighing lysimeters. *Proc. SPIE,* **5544,** 215–229.

Collins, W. D., and Coauthors, 2004: Description of the NCAR Community Atmosphere Model (CAM 3.0). NCAR Tech. Rep. NCAR/TN-464+STR, 214 pp. [Available online at http://www.cesm.ucar.edu/models/atm-cam/docs/description/description.pdf.]

Cook, D. R., D. J. Holdridge, and M. L. Fischer, 2006: Comparison of ECOR, EBBR, and CO2FLX4m system fluxes. *Proc. 16th ARM Science Team Meeting,* Albuquerque, NM, DOE/ARM.

Crow, W. T., and E. F. Wood, 2003: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.,* **26,** 137–149.

Derbyshire, S. H., 1995: Stable boundary layers: Observations, models and variability. Part I: Modeling and measurements. *Bound.-Layer Meteor.,* **74,** 19–54.

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

GAMDT, 2004: The new GFDL global atmosphere and land model AM2–LM2: Evaluation with prescribed SST simulations. *J. Climate,* **17,** 4641–4673.

Garratt, J. R., 1992: Extreme maximum land surface temperatures. *J. Appl. Meteor.,* **31,** 1096–1105.

——, and B. B. Hicks, 1973: Momentum, heat, and water vapor transfer to and from natural and artificial surfaces. *Quart. J. Roy. Meteor. Soc.,* **99,** 680–687.

Hartke, G. J., and D. Rind, 1997: Improved surface and boundary layer models for the Goddard Institute for Space Studies general circulation model. *J. Geophys. Res.,* **102,** 16 407–16 422.

Hill, R. J., 1997: Applicability of Kolmogorov's and Monin's equations of turbulence. *J. Fluid Mech.,* **353,** 67–81.

Högström, U., 1988: Non-dimensional wind and temperature profiles in the atmospheric surface layer: A re-evaluation. *Bound.-Layer Meteor.,* **42,** 55–78.

Hollinger, S. E., and C. S. T. Daughtry, 1999: Southern Great Plains 1997 Hydrology Experiment: Vegetation sampling and data documentation. Contract Rep. 644 to United States Department of Agriculture on Contract AG-58-1270-7-043, 36 pp.

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.

——, E. I. F. De Bruijn, and H.-L. Pan, 1990: A high-resolution air mass transformation model for short range weather forecasting. *Mon. Wea. Rev.,* **118,** 1561–1575.

Ingwersen, J., and Coauthors, 2011: Comparison of Noah simulations with eddy covariance and soil water measurements at a winter wheat stand. *Agric. For. Meteor.,* **151,** 345–355.

Janjić, Z. I., 1996: The surface layer in the NCEP Eta Model. Preprints, *11th Conf. on Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., 354–355.

Jiang, L., S. Islam, and T. N. Carlson, 2004: Uncertainties in latent heat flux measurement and estimation: Implications for using a simplified approach with remote sensing data. *Can. J. Remote Sens.,* **30,** 769–787.

Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, 2012: A revised scheme for the WRF surface layer formulation. *Mon. Wea. Rev.,* **140,** 898–918.

Kader, B. A., and A. M. Yaglom, 1990: Mean fields and fluctuation moments in unstably stratified turbulent boundary layers. *J. Fluid Mech.,* **212,** 637–662.

Kustas, W. P., B. J. Choudhury, M. S. Moran, R. J. Reginato, R. D. Jackson, L. W. Gay, and H. L. Weaver, 1989: Determination of sensible heat flux over sparse canopy using thermal infrared data. *Agric. For. Meteor.,* **44,** 197–216.

——, M. C. Anderson, J. M. Norman, and F. Li, 2007: Utility of radiometric–aerodynamic temperature relations for heat flux estimation. *Bound.-Layer Meteor.,* **122,** 167–187.

Lhomme, J.-P., B. Monteny, and M. Amadou, 1994: Estimating sensible heat flux from radiometric temperature over sparse millet. *Agric. For. Meteor.,* **68,** 77–91.

Lu, J., and M. Cai, 2009: Stabilization of the atmospheric boundary layer and the muted global hydrological cycle response to global warming. *J. Hydrometeor.,* **10,** 347–353.

Mahrt, L., 1998: Stratified atmospheric boundary layers and breakdown of models. *Theor. Comput. Fluid Dyn.,* **11,** 263–279.

——, 1999: Stratified atmospheric boundary layer. *Bound.-Layer Meteor.,* **90,** 375–396.

Merlin, O., and A. Chehbouni, 2004: Different approaches in estimating heat flux using dual angle observations of radiative surface temperature. *Int. J. Remote Sens.,* **25,** 275–289.

Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the atmosphere near the ground. *Akad. Nauk. U. S. S. R. Trud. Gefiz. Inst. Tr.,* **24,** 163–187.

Pahlow, M., M. B. Parlange, and F. Porté-Agel, 2001: On Monin–Obukhov similarity in the stable atmospheric boundary layer. *Bound.-Layer Meteor.,* **99,** 225–248.

Paulson, C. A., 1970: The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *J. Appl. Meteor.,* **9,** 857–861.

Pleim, J. E., 2006: A simple, efficient solution of flux–profile relationships in the atmospheric surface layer. *J. Appl. Meteor. Climatol.,* **45,** 341–347.

Santanello, J. A., M. A. Friedl, and M. B. Ek, 2007: Convective planetary boundary layer interactions with the land surface at diurnal time scales: Diagnostics and feedbacks. *J. Hydrometeor.,* **8,** 1082–1097.

Schmidt, G. A., and Coauthors, 2006: Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. *J. Climate,* **19,** 153–192.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech Note NCAR/TN-475+STR, 113 pp. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.]

Stokes, G. M., and S. E. Schwartz, 1994: The Atmospheric Radiation Measurement (ARM) program: Programmatic background and design of the cloud and radiation test bed. *Bull. Amer. Meteor. Soc.,* **75,** 1201–1221.

Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic, 666 pp.

Sun, J., 1999: Diurnal variations of thermal roughness height over a grassland. *Bound.-Layer Meteor.,* **92,** 407–427.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.,* **106,** 7183–7192.

Twine, T. E., W. P. Kustas, J. M. Norman, D. R. Cook, P. R. Houser, T. P. Meyers, J. H. Prueger, and P. J. Starks, 2000:

Correcting eddy covariance flux underestimates over a grassland. *Agric. For. Meteor.,* **103,** 279–300.

van den Hurk, B. J. J. M., F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne, and H. Camargo Jr., 2012: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Climate Dyn.,* **38,** 349–362.

Wu, W., Y. Liu, and A. K. Betts, 2012: Observationally-based evaluation of NWP reanalyses in modeling cloud properties over the southern Great Plains. *J. Geophys. Res.,* **117,** D12202, doi:10.1029/2011JD016971.

Yang, F., H.-L. Pan, S. K. Krueger, S. Moorthi, and S. J. Lord, 2006: Evaluation of the NCEP Global Forecast System at the ARM SGP Site. *Mon. Wea. Rev.,* **134,** 3668–3690.

Ye, Z., and X. Jia, 1991: The impact of soil moisture availability upon the partition of net radiation into sensible and latent heat fluxes. *Adv. Atmos. Sci.,* **8,** 339–350.

Zeng, X., and R. E. Dickinson, 1998: Effect of surface sublayer on surface skin temperature and fluxes. *J. Climate,* **11,** 537–550.

Zhan, X., W. Kustas, and K. Humes, 1996: An intercomparison study on models of sensible heat flux over partial canopy surfaces with remotely sensed surface temperature. *Remote Sens. Environ.,* **S8,** 242–256.

Zhang, D.-L., and R. A. Anthes, 1982: A high-resolution model of the planetary boundary layer—Sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.,* **21,** 1594–1609.

Zibognon, M., R. Crago, and A. Suleiman, 2002: Conversion of radiometric to aerodynamic surface temperature with an anisothermal canopy model. *Water Resour. Res.,* **38,** 1067, doi:10.1029/2001WR000484.

Zilitinkevich, S. S., 1970: *Dynamics of the Atmospheric Boundary Layer*. Leningrad Gidrometeor, 291 pp.

——, 1995: Non-local turbulent transport: pollution dispersion aspects of coherent structure of convective flows. *Air Pollution III—Volume I. Air Pollution Theory and Simulation,* H. Power, N. Moussiopoulos, and C.A. Brebbia, Eds., Computational Mechanics Publications, 53–60.