2-2015

Multiscale modeling and evaluation of urban surface energy balance in the Phoenix metropolitan area

Stephen R. SHAFFER

Winston T. L. CHOW
Singapore Management University, winstonchow@smu.edu.sg

Matei GEORGESCU

Peter HYDE

Darrel JENERETTE

See next page for additional authors

Citation

SHAFFER, Stephen R., CHOW, Winston T. L., GEORGESCU, Matei, HYDE, Peter, JENERETTE, Darrel, MAHALOV, Alex, MOUSTAOUI, M., & RUDDELL, Benjamin Lyle.(2015). Multiscale modeling and evaluation of urban surface energy balance in the Phoenix metropolitan area. *Journal of Applied Meteorology and Climatology, 54(2)*, 322-338.
Available at: https://ink.library.smu.edu.sg/soss_research/3058

Multiscale Modeling and Evaluation of Urban Surface Energy Balance in the Phoenix Metropolitan Area

S. R. SHAFFER,*,+ W. T. L. CHOW,# M. GEORGESCU,@ P. HYDE,+ G. D. JENERETTE,& A. MAHALOV,* M. MOUSTAOUI,* AND B. L. RUDDELL**

* *School of Mathematical and Statistical Sciences, Arizona State University, Tempe, Arizona*
+ School *for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, Arizona*
# Department *of Geography, National University of Singapore, Singapore*
@ School *of Geographical Sciences and Urban Planning, Arizona State University, Tempe, Arizona*
& Department *of Botany and Plant Sciences, University of California, Riverside, Riverside, California*
** *College of Technology and Innovation, Arizona State University, Tempe, Arizona*

(Manuscript received 5 March 2014, in final form 22 October 2014)

ABSTRACT

Physical mechanisms of incongruency between observations and Weather Research and Forecasting (WRF) Model predictions are examined. Limitations of evaluation are constrained by (i) parameterizations of model physics, (ii) parameterizations of input data, (iii) model resolution, and (iv) flux observation resolution. Observations from a new 22.1-m flux tower situated within a residential neighborhood in Phoenix, Arizona, are utilized to evaluate the ability of the urbanized WRF to resolve finescale surface energy balance (SEB) when using the urban classes derived from the 30-m-resolution National Land Cover Database. Modeled SEB response to a large seasonal variation of net radiation forcing was tested during synoptically quiescent periods of high pressure in winter 2011 and premonsoon summer 2012. Results are presented from simulations employing five nested domains down to 333-m horizontal resolution. A comparative analysis of model cases testing parameterization of physical processes was done using four configurations of urban parameterization for the bulk urban scheme versus three representations with the Urban Canopy Model (UCM) scheme, and also for two types of planetary boundary layer parameterization: the local Mellor–Yamada–Janjić scheme and the nonlocal Yonsei University scheme. Diurnal variation in SEB constituent fluxes is examined in relation to surface-layer stability and modeled diagnostic variables. Improvement is found when adapting UCM for Phoenix with reduced errors in the SEB components. Finer model resolution is seen to have insignificant (<1 standard deviation) influence on mean absolute percent difference of 30-min diurnal mean SEB terms.

1. Introduction

The aggregate global-scale impact of human activities is suggested to have brought about a geological epoch known as the Anthropocene (Smith and Zeder 2013). Most noticeable since the Industrial Revolution, anthropogenic influence may result in adverse transitions beyond critical thresholds, triggering ecosystem collapse (Barnosky et al. 2012). The world population is rapidly increasing and urbanizing while also increasing energy use and emissions (i.e., Ching 2013). The preeminent influence of anthropogenically determined local-scale urban microclimate is thus becoming ever more important within hot arid cities (e.g., Coutts et al. 2007). These cities are growing worldwide and are particularly vulnerable to climate change and water resource availability (i.e., Vörösmarty et al. 2010). These issues motivate the development of fine-resolution modeling tools for studying effects of urban design on a regional-scale to mitigate adverse effects and optimize urban microclimate. Modeled values of temperature and moisture provide key results to inform policy making and decisions regarding human–ecosystem interaction (Fernando 2008; Chow et al. 2012; Georgescu et al. 2013), although lack of available observations, particularly of surface energy balance (SEB) fluxes within urban settings, often leaves such predictions unvetted.

Corresponding author address: Stephen Shaffer, School of Mathematical and Statistical Sciences, Arizona State University, P.O. Box 871804, Tempe, AZ 85287-1804.
E-mail: stephen.shaffer@asu.edu

DOI: 10.1175/JAMC-D-14-0051.1

Furthermore, many features of urban microclimate are determined at scales < 1 km (Hunt et al. 2012; Ching 2013).

The capability of atmospheric modeling in urban environments is influenced by land–atmosphere coupling (Chen et al. 2011). The SEB is intimately related to the atmospheric surface layer (ASL), or region closest to the ground, which provides, via a surface-layer scheme (SLS), the interconnection between the ground, or land surface model (LSM), and lower atmosphere, or the planetary boundary layer (PBL). In particular, SEB closure has been examined in the context of urban climate and remains a challenging issue (Arnfield 2003; Foken 2008). In the context of SEB modeling, the role of vegetation, moisture, latent heat flux, and anthropogenic forcing are important areas of active research (Arnfield 2003; Ching 2013). Grimmond et al. (2010) conducted an extensive intermodel comparison of offline urban canopy models and found that not all models correctly account for SEB closure. A systematic evaluation of the modeled SEB is thus needed before addressing the above applications.

Parameterizations of urban processes within atmospheric models typically presume that the city is entirely subgrid to the ASL. This modeling assumption means that the built environment should be contained within the surface layer or first full model level (Grimmond et al. 2010). One main concern is that anthropogenic waste heat and momentum modifications are only supplied to the first model level. However, one method often employed within studies of the ASL or of lower PBL profiles is to add extra model levels near the ground. Hence, there is a trade-off between explicitly resolving fine structure in the ASL, especially within the urban boundary layer, also in conjunction with flow dominated by complex terrain (Fernando 2010). Furthermore, parameterizations of the ASL often employ the Monin–Obukhov similarity theory (MOST; Monin and Obukhov 1954), wherein horizontal homogeneity is assumed, meaning that individual buildings and land uses at subgrid scales are not explicitly resolved. This assumption can break down in settings with inhomogeneous land use and land cover (LULC). Microscale LULC variations may become important when comparing with neighborhood-scale flux-tower measurements (Foken 2008; Nordbo et al. 2013).

Forecasting at smaller scales has become computationally possible by advances in technology and in numerical technique, such as nesting (Skamarock and Klemp 2008). Yet theoretical issues regarding turbulence closure and parameterization of PBL eddies remain a challenge as resolution approaches the so-called terra incognita (Wyngaard 2004). At scales finer than the terra incognita ($\lesssim 1$ km), large-eddy simulations are typically employed (e.g., Moeng et al. 2007). Recently, model development efforts have been focused on enabling transiting of the terra incognita (Chen et al. 2011). Yet computational feasibility often limits applications to research. Furthermore, there has been limited assessment of model error at terra incognita scales.

In this study we explore SEB, computational feasibility, model stability, and sensitivity to parameterization, when nesting from global analysis data to a resolution of 333 m. This study addresses questions of model feasibility and accuracy within a hot arid city, probing the limits of current model physics parameterization schemes, computational capability, and input data, evaluated in a manner consistent with available observations of SEB. Model SEB terms were evaluated with flux-tower observations located in an arid urban residential neighborhood (Chow et al. 2014), assessing for a range of seasonal and diurnal input radiative forcing, and physically explaining errors. Multiple customized high-resolution urban LULC datasets were incorporated to evaluate parameterization of urban LULC appropriate to fine-resolution modeling. Two turbulence closure model PBL schemes, Yonsei University (YSU; Hong et al. 2006) and Mellor–Yamada–Janjić (MYJ; Janjić 2001), were investigated with a data and model combination probing the limitation of approaching terra incognita.

2. Methods

The influence of resolution is explored by employing spatially and temporally nested computational domains. Modified input parameterizations of LULC are employed to represent urbanization specific for Phoenix, Arizona, derived from observed data. A comparative analysis is then conducted between two types of model turbulence closure and four urban physical parameterization schemes for varied representations of the Phoenix urban canopy.

a. Flux-tower observations and study period

Micrometeorological data were obtained from an eddy flux tower installed in a residential west Phoenix neighborhood (33.484°N, 112.143°W). Observed values were postprocessed into half-hourly block-averaged (from 10 Hz) turbulent and radiative data, along with related temperature and three-dimensional wind data. The instruments were installed at 22.1 m above ground level. Further details can be obtained from Chow et al. (2014) regarding site characteristics, instruments employed, data quality, correction procedures, and calculation of flux source areas. Turbulent flux footprint lengths are approximately 0.5–1 km from unstable to

stable surface layers, and the radiative flux source area is ≈0.5 km in diameter.

A range of cloud-free dry-period SEB forcing conditions were examined by selecting time frames during winter and premonsoon summer with available SEB observations. These periods are the 60-h period from 23 December (winter 2011) and the 72-h period from 17 June (premonsoon summer 2012).

b. Numerical simulations

The Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008) was evaluated using observations described above. Simulations were conducted using one-way nested domains where multiple domains were run concurrently with no feedback to parent domains. The outer domain and soil moisture were initialized with final operational global analysis (FNL) data. These data are provided at 1° spatial and 6-h temporal resolution, at 27 vertical pressure levels. Radiative processes are represented by the RRTM scheme for longwave (Mlawer et al. 1997) and the Dudhia scheme for shortwave (Dudhia 1989). Physical processes involving moisture were modeled using the three-class single-moment microphysics scheme (Hong et al. 2004). The Kain–Fritsch cumulus parameterization (Kain 2004) was used for just the outer domain.

We utilize the Noah LSM described in Chen and Dudhia (2001), which determines skin temperature and supplies heat, momentum, and moisture fluxes into the atmosphere in response to radiation, precipitation, humidity, and surface-layer temperature and winds, for the dominant nonurban LULC. The geographic nonurban LULC classifications and terrain elevations were obtained from nearest-neighbor interpolation of the Moderate Resolution Imaging Spectroradiometer (MODIS) 20-category 30 arc s data modified for the Noah LSM. Vegetation fraction values were obtained from static terrestrial data provided in WRF.

1) Model resolution

Five nested domains, referred to as D_1–D_5, were configured with horizontal resolutions Δ_H of 27, 9, 3, 1, and 0.333 km, respectively, and are represented schematically in Fig. 1. Also shown are terrain for all domains and dominant LULC for D_5, which was configured to contain the entire Phoenix metropolitan area. All domains used a vertical grid with 40 vertical levels with increased grid resolution near the ground and a model top of 50 hPa. The first three domains were started synchronously. However, the starting times for D_4 and D_5 are delayed by 6 h each to allow for spinup. The D_1 time step for winter simulations was 150 s. Summer simulations required a reduction to as low as 90 s because of model stability. Attribution of reducing the child domain time steps by a factor of 3 from the parent domain for D_2–D_4, and by a factor of 5 for D_5, helped to stabilize the finescale domain simulation. The D_1 time step was further constrained to ensure that the D_2 time step would evenly divide the $\Delta_t = 300$ s output history interval.

2) Urban LULC parameterizations

The categorical urban LULC fields from the 30-m-resolution 2006 U.S. Geological Survey National Land Cover Database (NLCD; Fry et al. 2011) were used to derive representative urban LULC for the Phoenix metropolitan area. Three urban LULC classes were identified as commercial/industrial (C/I), high-intensity residential (HIR), and low-intensity residential (LIR). The C/I was derived from developed high-intensity, HIR was derived from developed medium-intensity, and LIR was derived from developed low-intensity and developed open space. Grid-scale urban LULC were then obtained by nearest-neighbor interpolation and made a higher priority when combining with the MODIS LULC classes. To obtain the final LULC product for each domain (Fig. 1), any grid cells still classified as urban/built-up by MODIS were replaced with LIR. Urban schemes are applied for the dominant urban LULC within each model grid cell to which an urban LULC was attributed.

We compare the role of urban parameterization for the bulk urban scheme (bulk), described in Liu et al. (2006), versus the urban canopy model (UCM) described in Kusaka and Kimura (2004). The bulk scheme calculates fluxes from a single flat surface. However, the UCM accounts for unresolved simplified infinite urban canyons, with building morphology and materials, roads, and interactions between roads, roofs, and walls of buildings, with 20 parameters for each urban LULC class.

We test three UCM cases by applying the calibrated LULC parameterizations. First, we test a baseline case (default), then we examine two other cases using modified morphological and material values for Phoenix (PHX-A and PHX-B; see Table 1). Only parameters that were changed from default values are given in Table 1, all of which remain fixed for PHX-A and PHX-B, except for the urban fraction f_{urb}, also given in Table 1 for all cases. The HIR class, which varies the most between PHX-A and PHX-B, also happens to be the LULC designation for the flux-tower footprint neighborhood. Thus, the differences between the three UCM cases will provide a simple variation of parameters.

The f_{urb} value represents the computational gridcell fraction attributed to the dominant urban LULC. For comparison, f_{urb} values for the three UCM cases are also

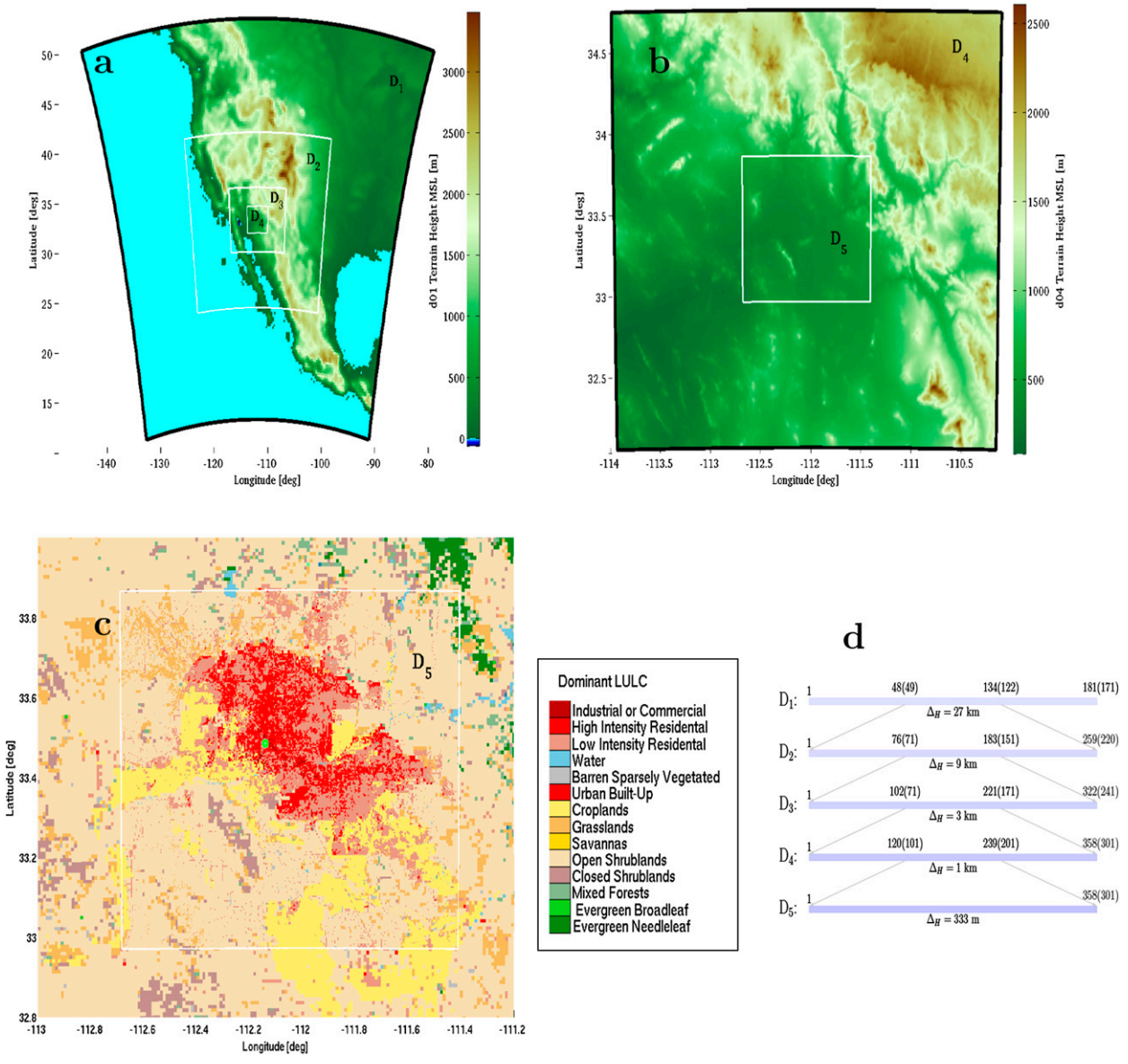

FIG. 1. WRF domain nest positions (with D_1 center at 33.5°N, 112°W) overlaid on terrain height above mean sea level (m) for (a) D_1–D_4 and (b) D_5 within D_4. (c) Dominant LULC given in legend (bottom middle) at grid resolution of D_4–D_5 and the west Phoenix flux-tower location (green circle; 33.484°N, 112.143°W). (d) A schematic indicating nesting by horizontal grid index, denoted by indices for east–west with north–south within parentheses, for D_1–D_5 with resolution Δ_H indicated.

presented in Table 1. The f_{urb} values are held constant for all domains for their respective urban LULC class when that class is the dominant LULC in a given model grid cell. Unlike the implementation of UCM, which employs three urban LULC classes with $f_{\text{urb}} \in (0, 1]$, the bulk scheme employs a single developed/built-up urban LULC class with $f_{\text{urb}} = 1.0$. Furthermore, f_{urb} is used as a coefficient of the UCM scheme output variables, with the nonurban variable (with coefficient $1 - f_{\text{urb}}$) derived from the Noah model. These subgrid fractional contributions are then aggregated to compute a single value for each grid cell.

The UCM scheme assumes that the built environment is subgrid to the first vertical model level. Our choice of 40 vertical levels provides for $z_1 \approx 55$ m for the first model-layer thickness, satisfying the subgrid condition for 99% of the buildings within the 16.7 km^2 core downtown Phoenix study area (Burian et al. 2002), where 73% are <5 m tall, and another 20% are between 5 and 10 m. While there are a few other built-up urban

TABLE 1. Description of modifications made to UCM urban physics option parameters (Kusaka and Kimura 2004) from default values following Grossman-Clarke et al. (2010) for urban LULC classes C/I, HIR, and LIR.

Parameter description	Units	C/I	HIR	LIR
Urban fraction[a] default values	—	0.865	0.429	0.429
Urban fraction[a] PHX-A values	—	0.95	0.60	0.73
Urban fraction[a] PHX-B values[b]	—	0.95	0.85	0.70
Roof level (building height)	m	10.0	4.7	3.9
Std dev of roof height[c]	m	8.0	2.7	1.0
Roof (i.e., building) width	m	31.7	25.7	17.6
Road width	m	98.9	39.2	108.0
Anthropogenic heat	$W\,m^{-2}$	30.0	35.0	20.0
Volumetric heat capacity of roof[d,e]	$MJ\,m^{-3}\,K^{-1}$	1.32	1.32	1.32
Volumetric heat capacity of building wall[f,e]	$MJ\,m^{-3}\,K^{-1}$	2.11	1.52	1.52
Volumetric heat capacity of ground (road)[e]	$MJ\,m^{-3}\,K^{-1}$	1.94	1.94	1.94
Thermal conductivity of roof[d,e]	$J\,m^{-1}\,s^{-1}\,K^{-1}$	0.83	0.83	0.83
Thermal conductivity of building wall[e]	$J\,m^{-1}\,s^{-1}\,K^{-1}$	1.51	0.19	0.19
Thermal conductivity of ground (road)[e]	$J\,m^{-1}\,s^{-1}\,K^{-1}$	0.75	0.75	0.75

[a] f_{urb}, fraction of the urban landscape that does not have natural vegetation.
[b] Modified f_{urb} following Georgescu et al. (2011).
[c] Grossman-Clarke et al. (2005).
[d] Assume brick roof.
[e] Value from Oke (1987).
[f] Assume concrete wall for C/I and dense wood for HIR and LIR.

cores within the greater Phoenix metropolitan area, the predominant LULC is <10 m tall residential.

3) PHYSICAL PARAMETERIZATIONS OF ATMOSPHERIC TURBULENCE

The role of the PBL scheme, which parameterizes vertical mixing processes of unresolved turbulent motion, was examined for two methods of turbulence closure. Each PBL scheme depends upon a specific SLS to connect the LSM to the first atmospheric model level, and hence will be influenced by feedback with the SEB. The nonlocal scheme of YSU (Hong et al. 2006), explicitly treats entrainment between the free atmosphere and top of the boundary layer, in addition to a nonlocal gradient flux term to account for large eddies. YSU couples with the MM5 SLS (Zhang and Anthes 1982). The higher-order local closure scheme of MYJ requires the eta SLS (Janjić 2001). Both SLSs employ MOST and assume a horizontally homogeneous and stationary constant flux layer. Horizontal subgrid mixing was achieved with a second-order diffusion parameterization and a Smagorinsky first-order closure scheme.

c. *Methods used for comparing observations and model simulations*

1) PHYSICAL METRICS USED FOR EVALUATION

For WRF Model evaluation, we use the following relations, with variables described in Table 2:

TABLE 2. Description of WRF output variables used for flux analysis in Eq. (1) or elsewhere.

Variable	Output name	Units	Description
ΣQ_H	ACHFX	$J\,m^{-2}$	Accumulated surface sensible heat flux
ΣQ_E	ACLHF	$J\,m^{-2}$	Accumulated surface latent heat flux
α	ALBEDO	—	Surface albedo
ϵ	EMISS	—	Surface emissivity
$G^{\downarrow}_{LW}$	GLW	$W\,m^{-2}$	Downward longwave flux at ground surface
Q_G	GRDFLX	$W\,m^{-2}$	Ground energy flux, positive release
Q_H	HFX	$W\,m^{-2}$	Surface sensible heat flux
Q_E	LH	$W\,m^{-2}$	Surface latent heat flux
z_{PBL}	PBLH	m	Boundary layer height
$G^{\downarrow}_{SW}$	SWDOWN	$W\,m^{-2}$	Downward shortwave flux at ground surface
T_{2m}	T2	K	2-m temperature
T_0	TSK	K	Surface skin temperature
u^*	UST	$m\,s^{-1}$	Friction velocity

$$
\begin{aligned}
G^{\uparrow}_{\mathrm{SW}} &= \alpha G^{\downarrow}_{\mathrm{SW}},\\
G^{\uparrow}_{\mathrm{LW}} &= \epsilon\sigma T_0^4,\\
Q^* &= G^{\downarrow}_{\mathrm{SW}} - G^{\uparrow}_{\mathrm{SW}} + G^{\downarrow}_{\mathrm{LW}} - G^{\uparrow}_{\mathrm{LW}}\\
\langle Q_H\rangle &= \Delta\Sigma Q_H/\Delta_t, \quad \text{and}\\
\langle Q_E\rangle &= \Delta\Sigma Q_E/\Delta_t. \qquad (1)
\end{aligned}
$$

Here $\sigma = 5.670\,40 \times 10^{-8}\,\mathrm{J\,s^{-1}\,m^{-2}\,K^{-4}}$ is the Stefan–Boltzmann constant, and Δ_t is the 5-min history output interval. Top to bottom are the upward radiative fluxes for shortwave $G^{\uparrow}_{\mathrm{SW}}$ and longwave $G^{\uparrow}_{\mathrm{LW}}$, and the net radiation Q^*. Last are fluxes of sensible $\langle Q_H\rangle$ and latent $\langle Q_E\rangle$ heat, for which accumulated quantities were used for comparison with time-averaged observations rather than instantaneous values. All quantities have units of $\mathrm{W\,m^{-2}}$.

The SEB relation for the effective residual or storage ΔQ_s is

$$\Delta Q_s = Q^* - \langle Q_H\rangle - \langle Q_E\rangle. \qquad (2)$$

When explicitly partitioning the anthropogenic forcing Q_F and ground heat flux Q_G, one could write $\Delta Q_s = \Delta Q'_s + Q_F + Q_G$, with a reduced residual $\Delta Q'_s$. However, for comparison with observations, Q_F is combined with the storage term ΔQ_s. The present analysis also considers Q_G as being a component of the residual term ΔQ_s because of the disparity in spatial scale between observation footprints of Q_G and the turbulent flux terms. Anthropogenic forcing Q_F was derived following Grossman-Clarke et al. (2005).

For making a fair comparison between observations and simulations, instantaneous values were output from WRF with $\Delta_t = 300\,\mathrm{s}$ and were averaged to the same 30-min periods. Percent differences P_D were calculated between 30-min diurnally averaged values for observation O and simulation S as

$$P_D = 100 \times \frac{O - S}{0.5(O + S)}, \qquad (3)$$

with variance estimated from standard deviation of averaged values and propagated to estimate statistical uncertainty in each 30-min interval of P_D. Furthermore, the scales of P_D, which can be both positive and negative and vary quite widely, are presented logarithmically in Figs. 2–7 (described later) as

$$\mathrm{Scale}(P_D) = \frac{P_D}{|P_D|} \times \log_{10}\lceil |10^s P_D| \rceil, \qquad (4)$$

where the coefficient will preserve the sign of P_D. We set the scaling factor $s = 0$ within the ceil function, $\lceil\cdot\rceil$, since we are interested in visually inspecting values of order of magnitude with $|P_D| > 10°$. Note that simulations more closely match observations with a smaller value of $|P_D|$.

2) Diagnostic temperature

The diurnal variation and percent differences between observations and model cases described in section 2b are shown for the diagnostic temperature at 2 m above ground $T_{2\mathrm{m}}$. Values of $T_{2\mathrm{m}}$ are calculated within WRF by the relation

$$T_{2\mathrm{m}} = T_0 - \frac{Q_H}{\rho c_p C_H}. \qquad (5)$$

Here Q_H and C_H are sensible heat flux and heat exchange coefficient, respectively, which are taken from the previous time step; T_0 is the skin temperature; ρ is the air density; and c_p is the specific heat at constant pressure. The heat exchange parameter is defined by $C_H = u^*\theta^*/(\Delta\theta)$, where $\Delta\theta = T_0 - T_{2\mathrm{m}}$. The friction velocity u^* and turbulent temperature scale θ^*, in turn, make use of MOST integrated stability functions for momentum and heat. Stability profiles are empirical relationships that act as fits to surface and first model level values consistent with gradient flux relationships [see section 2b(3)], and so are influenced by model bias at both levels. In particular, a bias in any of the terms contributing to $T_{2\mathrm{m}}$ could lead to error in derived values, and in some cases can cancel yielding a derived $T_{2\mathrm{m}}$ that may be more accurate than the individual parameters from which it was obtained.

3) Model evaluation across resolution and parameterization configurations

A metric of model error is needed to enable intercomparison between the different cases of model horizontal grid resolution Δ_H and model parameterization configuration p. For this purpose, we examine the diurnal mean absolute percent difference between observations and simulations M, defined for a variable x by

$$M(x;\Delta_H,p) = N_k^{-1}\sum_{k=1}^{N_k} |P_D[x(k;\Delta_H,p)]|. \qquad (6)$$

The sum is over the $k = 1, \ldots, N_k = 48$ thirty-minute time intervals in the diurnal period, and P_D is as given by Eq. (3), for the model grid point containing the observation location.

3. Results and discussion

Here we address the main research questions pertaining to how well current WRF performs in a hot dry city (Phoenix) and where further improvement is

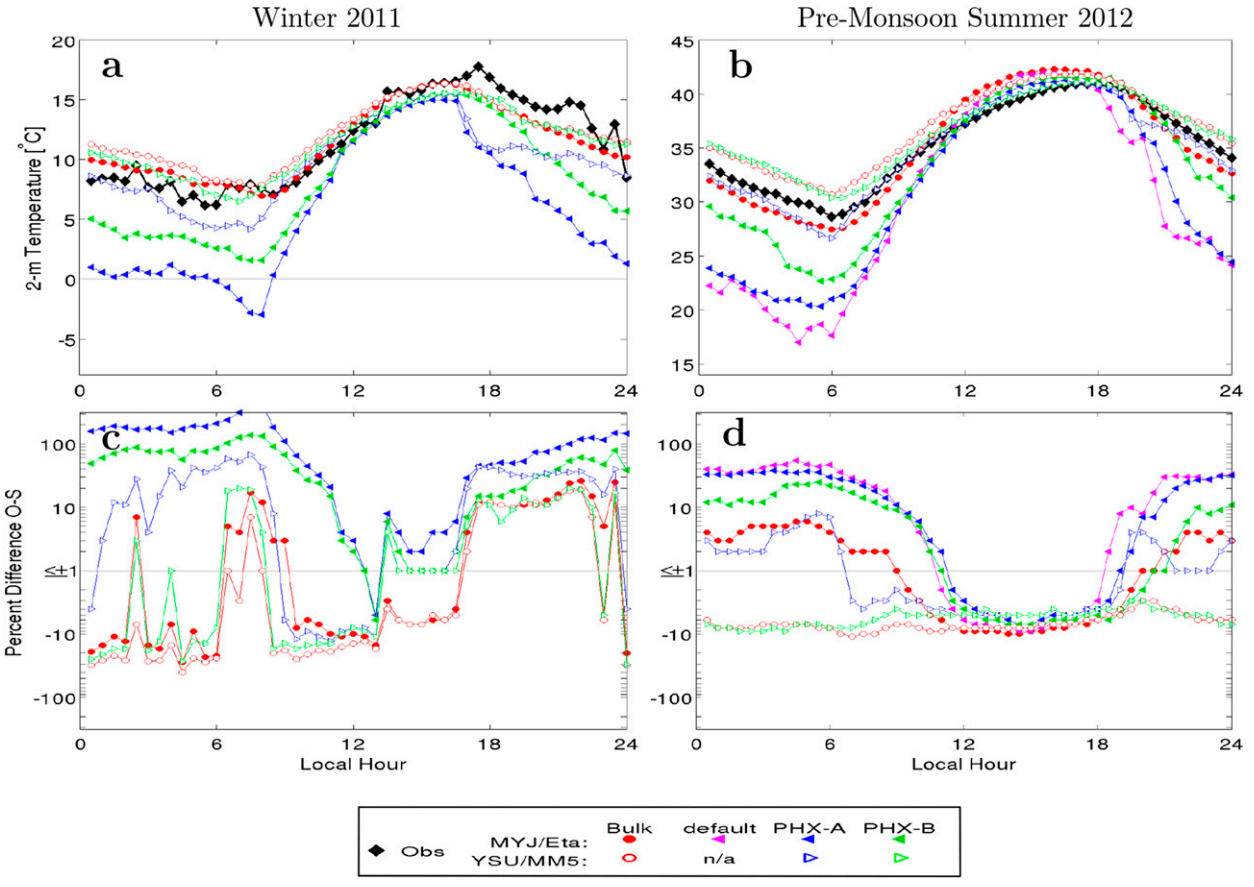

FIG. 2. Comparison of (a),(b) surface-layer 2-m temperature T_{2m}, between observed and simulated values, for D_4 ($\Delta_H = 1$ km) as 30-min and diurnally averaged values, and (c),(d) percent differences of simulations from observations for (left) winter 2011 and (right) pre-monsoon summer 2012. Shown are observed values (black diamonds), PBL-SLS for MYJ-eta (solid symbols), and YSU-MM5 (open symbols), with bulk urban scheme (red circles), and UCM urban scheme (triangles) for configurations of default (magenta; premonsoon summer 2012 MYJ-eta only), PHX-A (blue), and PHX-B (green), summarized in the legend (bottom). Note that percent-difference ordinate scale is limited to ±120% and has been made logarithmic preserving sign following Eq. (4).

needed, as validated with observations of SEB components. Inspection of 30-min averaged time series of observed values (not shown) reveals a regular diurnal quality for the chosen study period, wherein local flow processes are dominant over mesoscale forcing, which justifies the use of diurnal averages. The diurnal maxima of net radiative forcing vary by a factor of nearly 2 between the seasons.

A further distinction between the three default urban LULC classes and those employed by Grossman-Clarke et al. (2010) is that the latter values are representative of commercial–industrial, mesic residential, and xeric residential, respectively. Furthermore, soil moisture values were initialized in our cases from NCEP FNL data with no prescription for modifications to incorporate effects of irrigation, as conducted with the previous studies of Grossman-Clarke et al. (2010) and Georgescu et al. (2011). Rather, the present analysis is focused on examining the importance of f_{urb} values along with adapting the urban morphological and material parameters for Phoenix. The flux-tower footprint is a neighborhood with little vegetation and irrigation, and few water bodies (e.g., swimming pools; Chow et al. 2014). Furthermore, the NLCD data were from 2006, with negligible LULC modifications in the footprint area at the time of our study and more recent Quickbird-derived LULC (Chow et al. 2014, their Fig. 2 and Table 1).

The diurnally averaged variables simulated by the 1-km-resolution domain for WRF Model configurations defined in section 2b are compared with observed temperature and corresponding percent difference diurnal variation in (Fig. 2). Likewise, the radiation flux components are presented for $G^{\downarrow}_{LW}$ (Fig. 3), $G^{\uparrow}_{SW}$ (Fig. 4), and $G^{\uparrow}_{LW}$ (Fig. 5), and the friction velocity u^* is presented (Fig. 6). Similarly, the SEB flux quantities for just premonsoon summer 2012 are presented for $\langle Q_H \rangle$ and

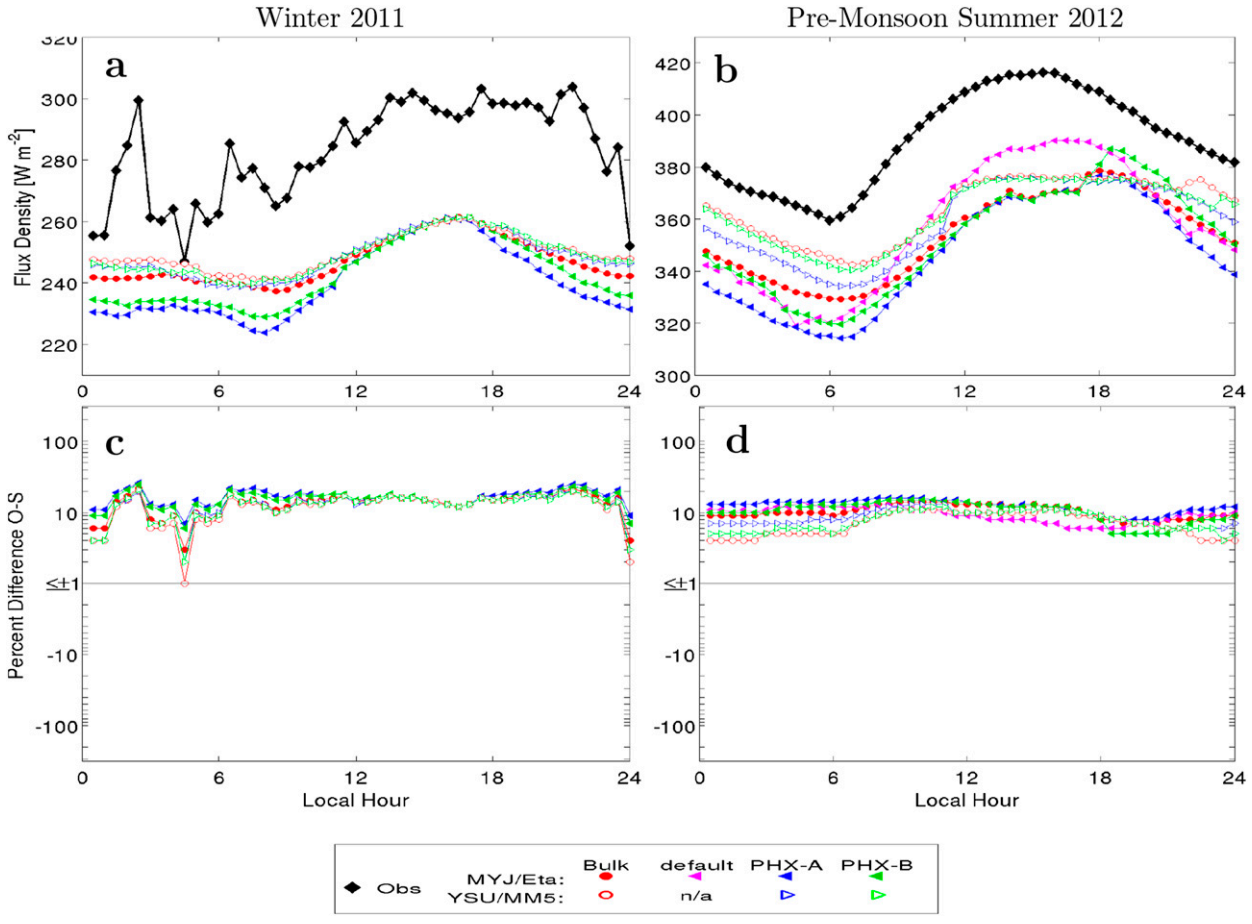

FIG. 3. As in Fig. 2, but for $G^{\downarrow}_{\mathrm{LW}}$.

$\langle Q_E \rangle$ (Fig. 7), and for Q^* and ΔQ_s (Fig. 8). Influence of resolution and configuration on model errors derived with Eq. (6) are presented in Fig. 9 for premonsoon summer 2012 for variables Q^*, $\langle Q_H \rangle$, $\langle Q_E \rangle$, ΔQ_s, $T_{2\mathrm{m}}$, $G^{\uparrow}_{\mathrm{LW}}$, $G^{\downarrow}_{\mathrm{LW}}$, $G^{\uparrow}_{\mathrm{SW}}$, and $G^{\downarrow}_{\mathrm{SW}}$, for $\Delta_H = 9$, 3, and 1 km, and for model configurations of bulk, PHX-A, and PHX-B, with either MYJ-eta or YSU-MM5. Results for the 333-m-resolution domain are only presented for model configurations MYJ-eta bulk, YSU-MM5 PHX-A, and YSU-MM5 PHX-B for summer 2012 (Fig. 9) because of the excessive computational time needed at this resolution.

a. *Influence of modifying turbulence parameterization, urban LULC, and urban representation input data*

The influence of local versus nonlocal closure schemes on bias for $T_{2\mathrm{m}}$ is apparent in Fig. 2. All cases perform well during midday unstable conditions. However, only cases MYJ-eta with bulk and YSU-MM5 with UCM for PHX-A or -B perform well at all times of day and for both seasons. The notation "PHX-A/B" will be used to represent UCM for either PHX-A or PHX-B. Also, the YSU-MM5 cases are warmer at night than corresponding MYJ-eta cases (Figs. 2a,b), and with a higher z_{PBL} and lower stability (not shown), agreeing with previous studies comparing local and nonlocal schemes (e.g., Xie et al. 2012). The role of f_{urb} is present at night (Figs. 2a,b), with PHX-B consistently warmer than PHX-A, for a given PBL-SLS. Here, the UCM scheme with MYJ-eta reduces $T_{2\mathrm{m}}$ underestimation error by ≈50%. For instance, (Fig. 2a) shows that PHX-A MYJ-eta underestimates $T_{2\mathrm{m}}$ by ≈8°C between 0000 and 0600 local time, while PHX-B MYJ-eta underestimated $T_{2\mathrm{m}}$ by ≈4°C during the same period. The bulk scheme evaluates well compared to the UCM scheme with regard to $T_{2\mathrm{m}}$, and YSU-MM5 for both PHX-B and bulk show quite similar $T_{2\mathrm{m}}$ for both seasons. However, this performance for bulk (Fig. 2) does not persist for the SEB terms (Figs. 7a,c), and for, for example, $G^{\uparrow}_{\mathrm{LW}}$ (Fig. 5a) and u^* (Figs. 6a,b) from which $T_{2\mathrm{m}}$ is derived [Eq. (5)].

A PBL-SLS dependence is also present for $G^{\downarrow}_{\mathrm{LW}}$ (Fig. 3), which may be due to feedback from the surface

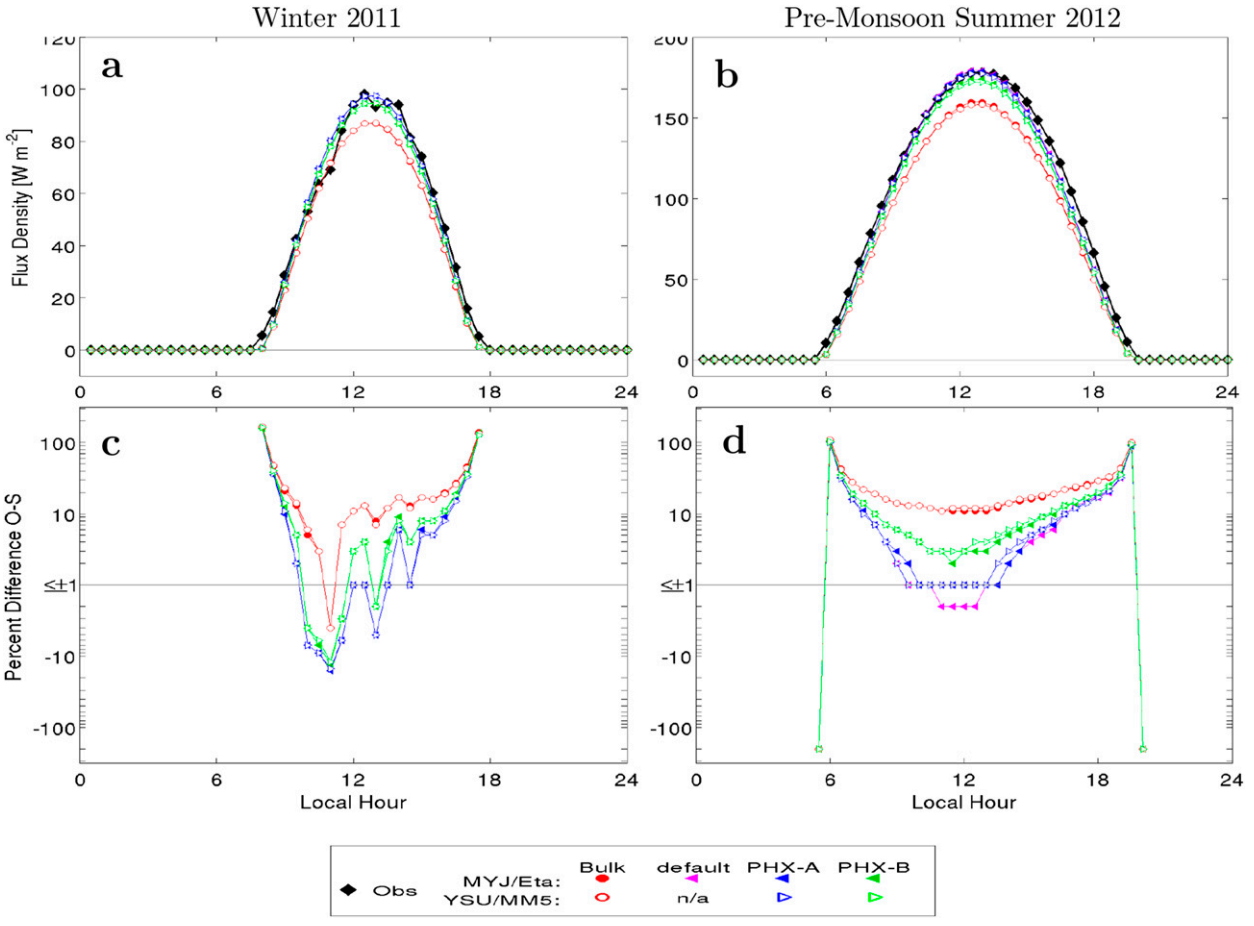

FIG. 4. As in Fig. 2, but for $G^{\uparrow}_{SW}$.

modifying the column temperature profile where the atmosphere is too cold. This explanation would agree with previous studies that have shown an influence of PBL-SLS between local and nonlocal schemes on derived temperature profiles (Shin and Hong 2011) and on SLS-LSM coupling strength (Chen and Zhang 2009). The systematic $G^{\downarrow}_{LW}$ underprediction bias (Figs. 3c,d) could also be related to the afternoon 3%–11% overprediction bias in $G^{\downarrow}_{SW}$ (not shown), and the lack of accounting for urban air pollutants, and a repartitioning of $G^{\downarrow}_{SW}$ into $G^{\downarrow}_{LW}$ by photochemically active species. These issues will be addressed in a future paper, and may further explain the $G^{\downarrow}_{LW}$ bias between models that use the same Dudhia and RRTM radiation schemes. The influence of bias in $G^{\downarrow}_{SW}$ is present in $G^{\uparrow}_{SW}$ (Fig. 4) owing to Eq. (1) and does not show a dependence upon PBL-SLS.

For $G^{\uparrow}_{SW}$ (Fig. 4), the cases with UCM have smaller error than bulk during midday because of differences in α and f_{urb}. For example, the summer afternoon $G^{\uparrow}_{SW}$ bias magnitude has values of $\gtrapprox$10% for bulk, ≈3% for PHX-B, ≈1% for PHX-A, and ≈−2% for default (Fig. 4d). The bulk scheme has $f_{urb} = 1$, but no accounting for buildings, roads, or other surfaces as with UCM, and so only the urban and built-up LULC where $\alpha = 0.15$ is attributed. Meanwhile, the three UCM cases use input values of α for road, roof, and wall, unlike the bulk scheme. The Noah LSM then accounts for f_{urb} and nonurban fraction $(1 - f_{urb})$ contributions to α. Thus differences between PHX-A and PHX-B are due solely to f_{urb} since α is identical for these two cases. Differences between default and the PHX-A/B cases are due to constituent values of α, along with building size and road width.

The modeled $G^{\uparrow}_{LW}$ has an afternoon underprediction bias near 10% (Figs. 5c,d). Inter-parameterization bias difference is reduced the most during the summer afternoon unstable period, with all cases underpredicting observations by ≈10% (Fig. 5d). The P_D of all cases, where P_D uses Eq. (3), also increases during winter afternoon unstable periods (Fig. 5c). The bulk scheme has $P_D \approx 10\%$ all day for both seasons and for both PBL-SLS cases. However, the $|P_D|$ of YSU-MM5 tends to be

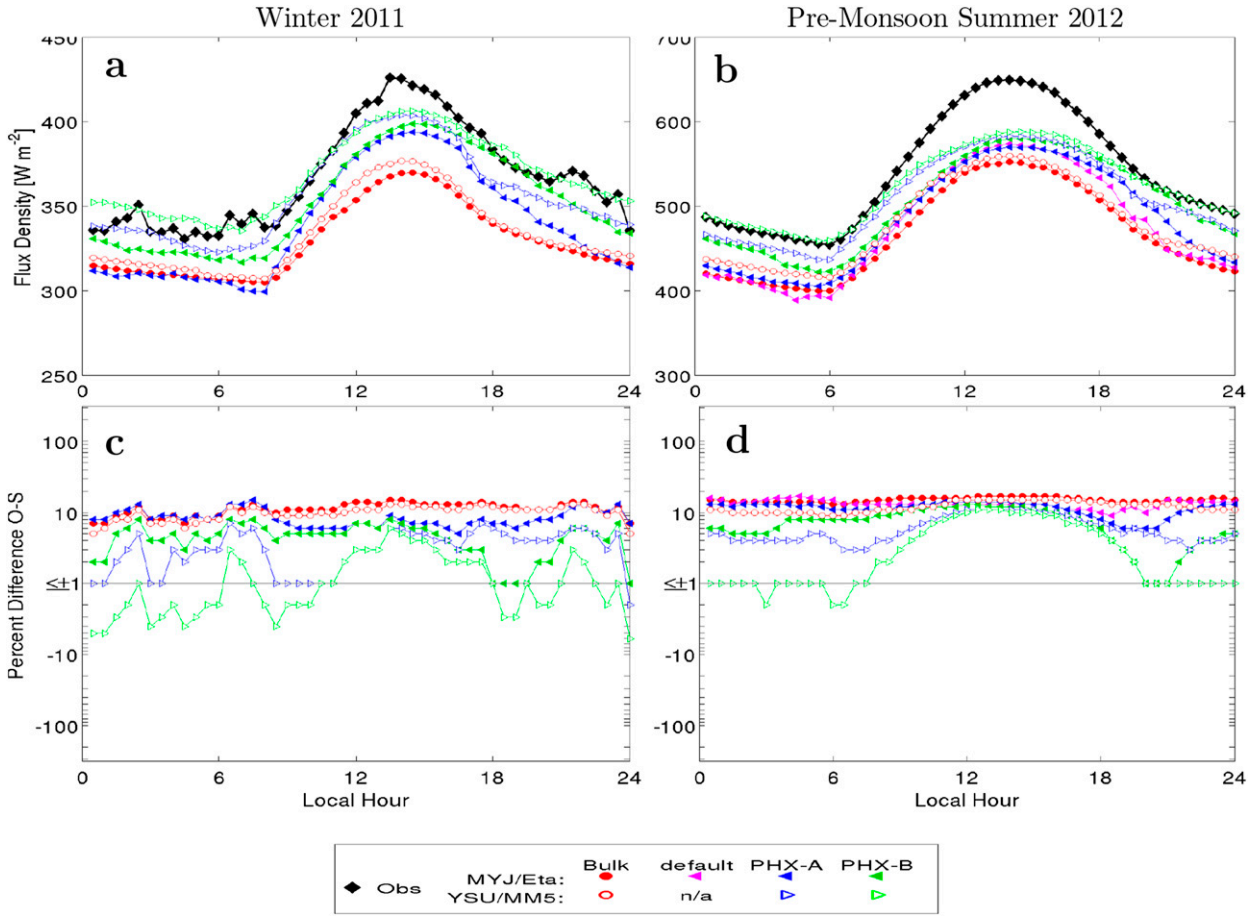

FIG. 5. As in Fig. 2, but for $G^{\uparrow}_{\mathrm{LW}}$.

smaller than MYJ-eta overnight, which remains ≈10%. For instance, between 0000 and 0600 local time for the summer PHX-B cases, the YSU-MM5 case has $|P_D| \leq 2\%$, whereas the MYJ-eta case has $P_D \geq 5\%$ (Fig. 5d). The reduced nocturnal P_D is present for both seasons (Fig. 5c,d) and is more pronounced for the winter period (Fig. 5c), which has longer stable conditions than during the summer period (not shown).

Evaluation of terms in Eq. (5) can reveal issues with model bias that are masked by bias cancelation within $T_{2\mathrm{m}}$. Here T_0 is examined with $G^{\uparrow}_{\mathrm{LW}}$ (Fig. 5), as per Eq. (1). Bias of $G^{\uparrow}_{\mathrm{LW}}$ depends on f_{urb} in nocturnal periods. Also, bias in $G^{\uparrow}_{\mathrm{LW}}$ has a PBL-SLS dependence, with either the role of f_{urb} reduced during the day or the role of PBL-SLS stability class becoming a factor. There are several feedbacks for T_0, including stability profiles $\langle Q_H \rangle$ and C_H. These variables are an aggregate of the UCM representation of the underlying urban fabric and the Noah contribution.

A sensitivity to f_{urb} is apparent within $G^{\uparrow}_{\mathrm{LW}}$ (Fig. 5), which, based upon [Eq. (1)], is most sensitive to T_0, and otherwise first order to f_{urb} through flux aggregation. For this particular neighborhood, the PHX-A/B shows improved results over the default case (Figs. 5c,d). The bulk scheme yields a systematic 10%–20% bias below observations that persists all day and for both seasons. Employing UCM for Phoenix reduces error in comparison with bulk, although not as significantly for premonsoon summer 2012 midday periods (Figs. 5b,d). The tendency in terms of $|P_D|$ is that PHX-B ≤ PHX-A ≤ bulk, for a given PBL-SLS case, except during midday summer where differences in mean values are less than variances of the means (not shown). Thus, at night, the higher f_{urb} PHX-B is warmer than PHX-A, seen both for T_0 via $G^{\uparrow}_{\mathrm{LW}}$ (Fig. 5) and for $T_{2\mathrm{m}}$ (Fig. 2).

The previously discussed $G^{\downarrow}_{\mathrm{SW}}$ bias (not shown) should lead to an overpredicted T_0, but $G^{\uparrow}_{\mathrm{LW}}$ has an underprediction bias (Fig. 5). By examination of $\langle Q_H \rangle$ during nocturnal premonsoon summer 2012 (Figs. 7a,b), modeled values employing YSU-MM5 deviate from the MYJ-eta cases, with the YSU-MM5 cases tending to have increased $\langle Q_H \rangle$ relative to both observations and

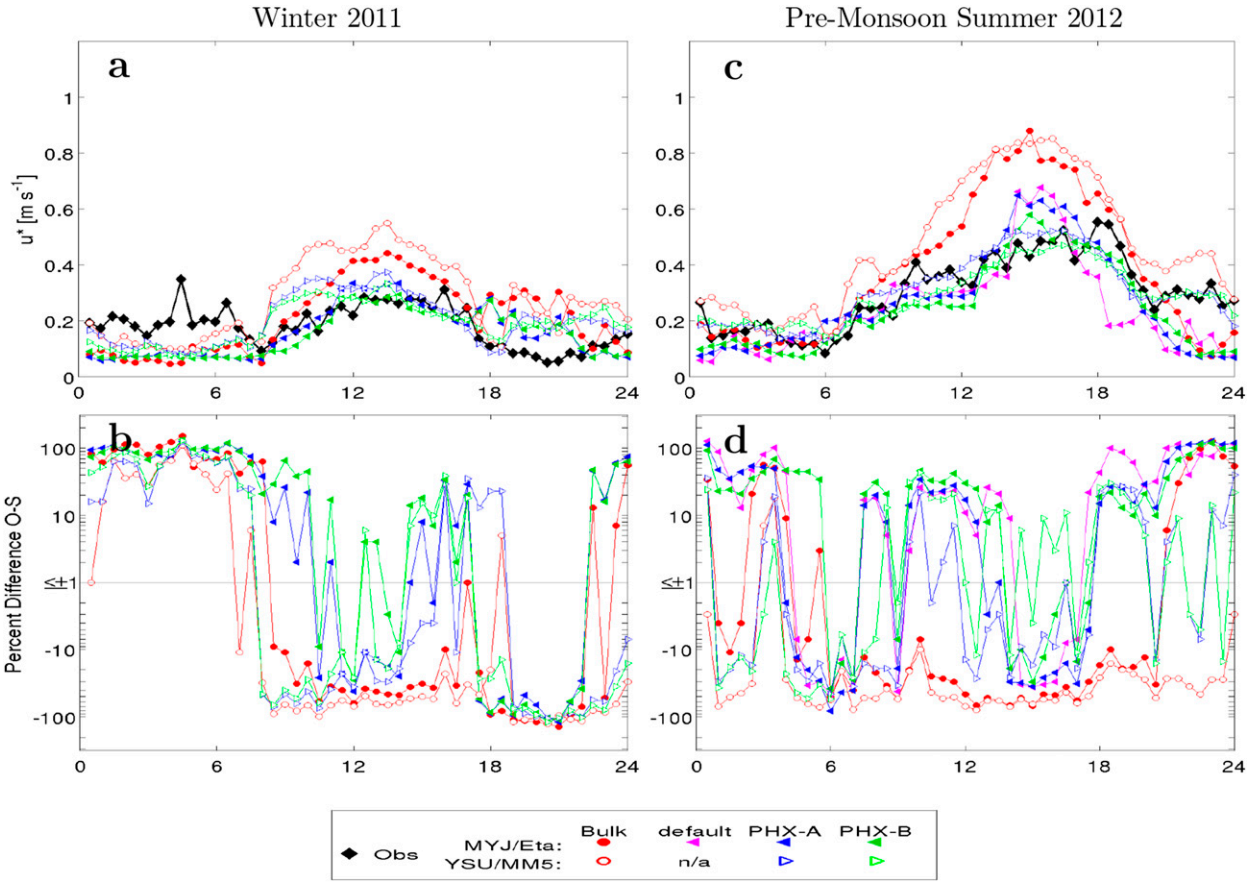

FIG. 6. As in Fig. 2, but for u^*.

the MYJ-eta cases. The inter–PBL-SLS case difference in T_0 for a fixed urban case (Fig. 5) must arise in T_{2m} owing to Eq. (5). However, $\langle Q_H \rangle$ (Figs. 7a,b) also indicates overestimated daytime heating for the larger f_{urb} case PHX-B, and an overestimation of $\langle Q_E \rangle$ (Figs. 7c,d) arises from the lower f_{urb} case PHX-A. These results indicate that the UCM scheme is missing physical mechanisms by which T_0 was increased (with coefficient f_{urb}), or the bias is arising from the nonurban "natural" contribution. One approach to reduce T_0 bias is to explore models beyond UCM that incorporate other afternoon processes that would increase $G^{\uparrow}_{LW}$ by modifying the effective T_0. Alternately, Wang et al. (2011) demonstrated that T_0 is most sensitive to uncertainties in urban geometry, thermal properties of surface materials, and roughness lengths, which suggests other values in Table 1 may lead to systematic bias.

The u^* term within C_H is another potential source of bias for T_{2m} and compares well to observations (Fig. 6). The UCM cases have lower u^* than for bulk because of increased roughness length. Afternoon magnitude of u^* for PHX-B is less than u^* for PHX-A because of f_{urb}. Both PHX-A/B cases were closer to observations than bulk during daytime. Especially during summer between 0900 and 1800 local time where P_D indicates bulk is typically overestimated by 30%–80%, and, for example, YSU $|P_D|$ is often under 30% (Figs. 6b,d). However, nocturnal periods were poorly represented by all cases as observations show intermittent increases in u^*, as in simulations, but a relation between events cannot be determined without more detailed observations. Stable nocturnal period dynamics are notoriously poorly modeled, a shortcoming that is suspected to be contributing to these errors in u^*.

For the $\langle Q_H \rangle$ term as a component of bias in T_{2m} [Eq. (5)], stable nocturnal periods show a bias dependent upon PBL-SLS for $\langle Q_H \rangle$ (Figs. 7a,b), which is much larger than the f_{urb} influence on $\langle Q_H \rangle$. Here, YSU-MM5 has an increased bias, typically underpredicting observed $\langle Q_H \rangle$, compared to MYJ-eta, which often overpredicted observations. Negative $\langle Q_H \rangle$ values indicate a release of heat from the surface, and so the bias due to PBL-SLS would lead to a warmer T_{2m} for YSU-MM5 than for MYJ-eta. These PBL-SLS–dependent differences in $\langle Q_H \rangle$ reduce

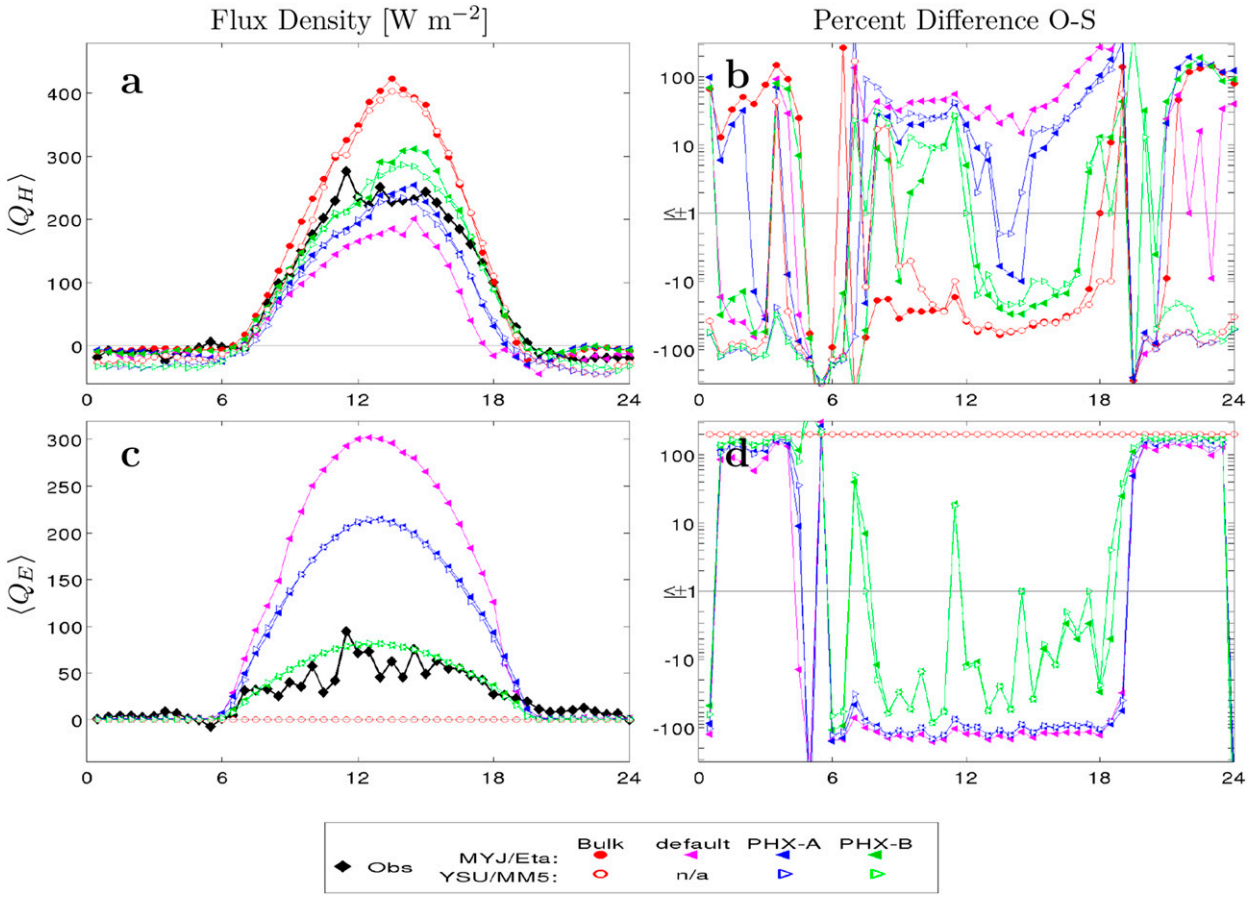

FIG. 7. As in Fig. 2, but only for during premonsoon summer 2012 and for (a),(c) 30-min and diurnally averaged observed and simulated values, and (b),(d) percent difference of simulations from observations, for (top) $\langle Q_H \rangle$ and (bottom) $\langle Q_E \rangle$.

during midday unstable periods, where f_{urb} becomes the dominant parameter for $\langle Q_H \rangle$ (Figs. 7a,b), and UCM case differences are low for T_{2m} (Fig. 2).

Figure 7a shows the $\langle Q_H \rangle$ diurnal cycle dependence on both urban land surface representation and PBL-SLS, with increased divergence of P_D between PBL-SLS cases during stable nocturnal periods with YSU-MM5 underpredicting observations (Fig. 7b). During the afternoon period, both PHX-A/B cases show close agreement with observations, with the lower f_{urb} PHX-A having better agreement during late afternoon, while both yield comparable predictions at night. The late-afternoon difference can be understood by the larger f_{urb} leading to increased heat transfer from the atmosphere to the urban canopy. A similar argument based upon f_{urb} applies to the default UCM case, which underpredicts daytime observations, and the bulk case, which overpredicts daytime observations.

The role of f_{urb} is apparent in $\langle Q_E \rangle$ bias (Fig. 7c) where moisture arises from the ''natural'' (i.e., nonurban) LULC via the Noah scheme. This explanation is clearly demonstrated by the bulk scheme reporting $\langle Q_E \rangle = 0$, since $1 - f_{urb} = 0$. There is no significant difference between the two PBL-SLS schemes noted for $\langle Q_E \rangle$ or $\langle Q_H \rangle$. Moreover, the agreement with observations for $\langle Q_E \rangle$ indicate that the larger f_{urb} (lower nonurban fraction) PHX-B gives better results than PHX-A, which is the opposite of what was concluded by examining $\langle Q_H \rangle$. This contradiction suggests that the contribution of vegetation is being overrepresented within the nonurban component and yields a point of caution encountered in modeling arid cities. Furthermore, the soil moisture was initialized to be too low (0.1 m^3 m^{-3} in top layer) relative to observed values (lowest value long after rain events approaches 0.1 m^3 m^{-3}; Chow et al. 2014), and no irrigation was applied.

During daytime and nocturnal periods, Q^* exhibits mixed bias (Figs. 8a,b), and it is also dependent upon which PBL-SLS and urban representation were employed. Similar results are present in the winter period simulations but with a shift that reduces midday overprediction bias and increasing nocturnal underprediction

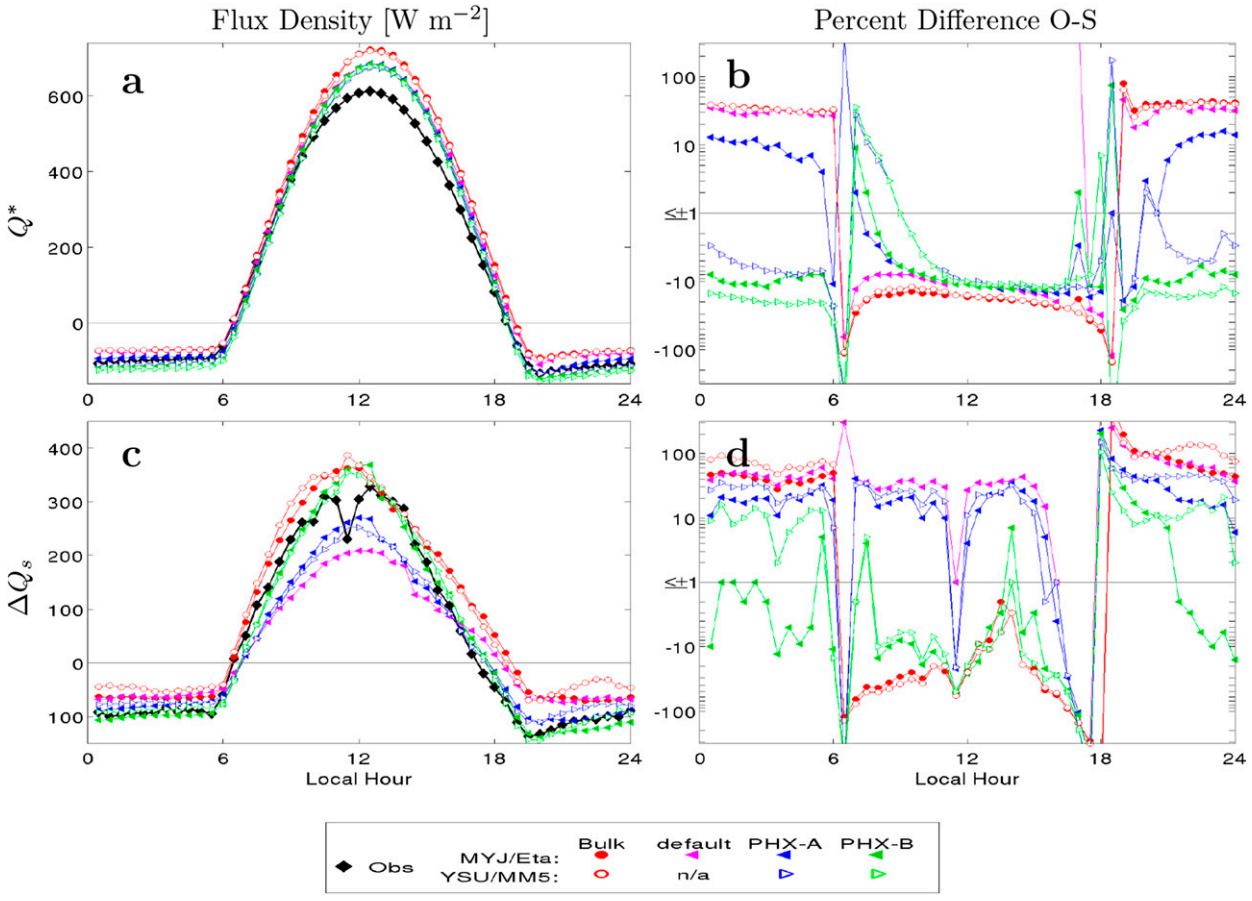

FIG. 8. As in Fig. 7, but for (a),(b) Q^* and (c),(d) ΔQ_s.

bias (not shown). Bias error cancelation influences the evaluation of variables composed of differences between terms. This is seen for cases and times where the bias errors in Q^* are comparable to or smaller than the bias in the individual radiative flux terms. For instance, at 2330 local time (since $G^{\uparrow}_{\mathrm{SW}} = G^{\downarrow}_{\mathrm{SW}} = 0$ at night), the MYJ-eta PHX-B case has $P_D = 4\%$ for $G^{\uparrow}_{\mathrm{LW}}$ (Fig. 5d), P_D for $G^{\downarrow}_{\mathrm{LW}}$ is 6% (Fig. 3d), and yet P_D for Q^* is only -2% (Fig. 8b). Thus, assessing Q^* in the absence of considering the constituent radiative flux components will not provide a robust model evaluation. Furthermore, it is difficult to disentangle potential sources of model error within the radiation forcing by solely examining bias in Q^*. For instance, an overestimation bias in solar forcing by the $G^{\downarrow}_{\mathrm{SW}}$ term and an underestimation bias in radiative cooling by the $G^{\uparrow}_{\mathrm{LW}}$ term may both lead to an overestimation of energy at the surface, and they may be related to an overestimation of convective cooling by the $\langle Q_H \rangle$ and $\langle Q_E \rangle$ terms.

The combination of terms in Eq. (2) used to derive ΔQ_s result in mixed performance between urban representation (Figs. 8c,d). The PHX-B configuration is more consistent all day than other cases, with $|P_D|$ often under 20%, with bulk being furthest from observations. The ordering of modeled ΔQ_s bias correlates with the f_{urb}. Afternoon observed ΔQ_s values are between PHX-A and PHX-B, indicating that an intermediate f_{urb} should more closely reproduce ΔQ_s during this period. Also, PHX-A with lower f_{urb} value often had lower $|P_D|$ during nighttime compared to PHX-B.

Choice of PBL-SLS has small significance in ΔQ_s (Figs. 8c,d), as $<1\sigma$ variations (not shown) for PBL-SLS cases arose within the $\langle Q_H \rangle$ and Q^* terms. The bulk case with MYJ-eta was seen to reduce errors in the predawn period compared to YSU-MM5, but still has $P_D > 20\%$–40%. Partial bias canceling is obscuring the $G^{\uparrow}_{\mathrm{LW}}$ and $G^{\downarrow}_{\mathrm{LW}}$ bias contributions to error within Q^*. Bias canceling is more complex within ΔQ_s, which is also influenced by $\langle Q_H \rangle$ and $\langle Q_E \rangle$ terms in Eq. (2). Hence, assessing ΔQ_s, along with SEB closure, can be challenging without exploring bias in all terms. Similar conclusions are drawn for SEB terms $\langle Q_H \rangle$, $\langle Q_E \rangle$, ΔQ_s, and Q^* during winter 2011 (not shown).

Figure 9 shows the influence of model parameterization on mean percent difference model bias error metric

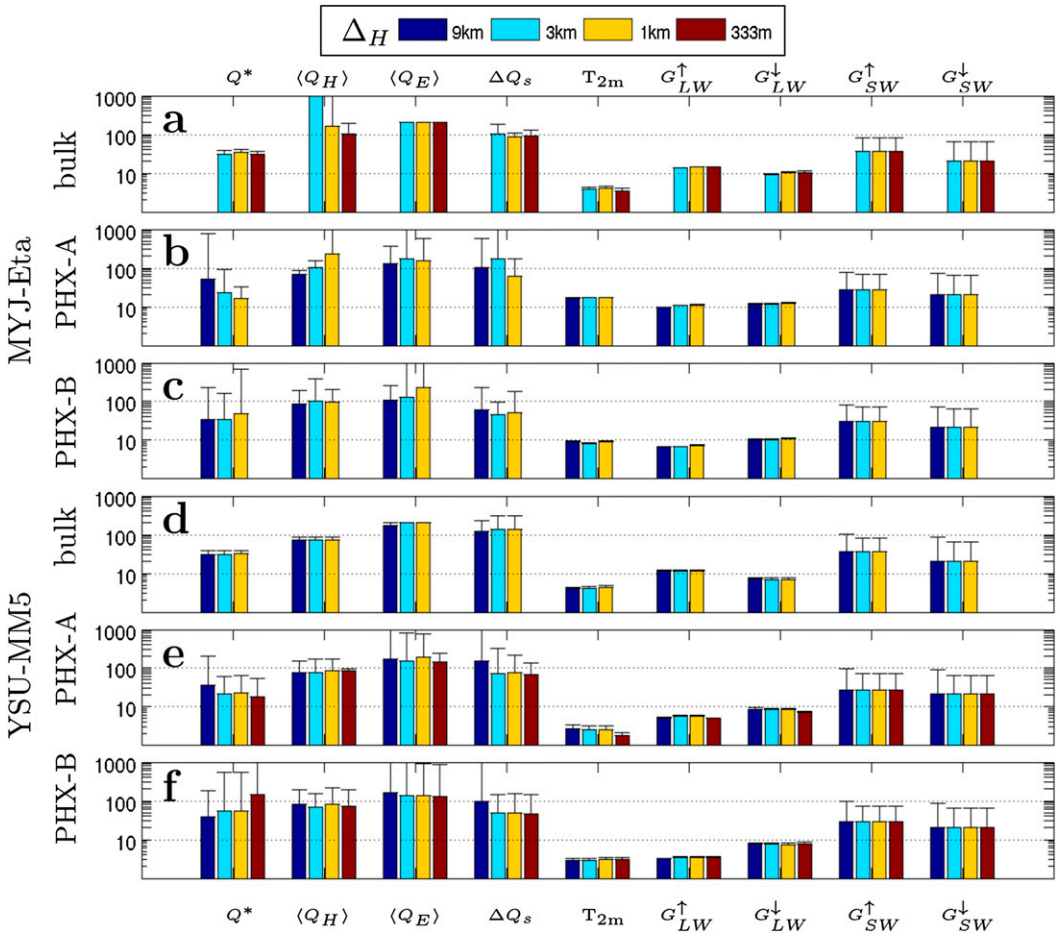

FIG. 9. Error metric M [Eq. (6)] as diurnal mean absolute percent difference between observations and simulations, for premonsoon summer 2012, for horizontal grid resolutions Δ_H, of 9, 3, and 1 km, and 333 m when available, for model configurations (a) MYJ-eta bulk, (b) MYJ-eta PHX-A, (c) MYJ-eta PHX-B, (d) YSU-MM5 bulk, (e) YSU-MM5 PHX-A, and (f) YSU-MM5 PHX-B, and for variables (left to right) Q^*, $\langle Q_H \rangle$, $\langle Q_E \rangle$, ΔQ_s, T_{2m}, $G^{\uparrow}_{LW}$, $G^{\downarrow}_{LW}$ $G^{\downarrow}_{LW}$, $G^{\uparrow}_{SW}$, and $G^{\downarrow}_{SW}$ for premonsoon summer 2012. Note that 9-km MYJ bulk output was only reported hourly and is excluded from this analysis.

$M(\cdot)$ for the entire diurnal cycle derived using Eq. (6) along with standard deviation error bars (with value of 1σ) for the 30-min diurnal mean P_D values. From these data, the effect of Δ_H on the error metric $M(\cdot)$ given by Eq. (6) is small relative to the σ. These figures demonstrate the influence of bias canceling between terms that compose T_{2m} in Eq. (5), with variables T_0 and $\langle Q_H \rangle$, for which we have already examined diurnal P_D at 1-km resolution. Examining Fig. 9, $M(T_{2m})$ is typically similar for $M(G^{\uparrow}_{LW})$ ($\lesssim$10%), whereas it is one order of magnitude less than $M(\langle Q_H \rangle)$. Bias of other terms from which T_{2m} is derived (not shown) would modify the $\langle Q_H \rangle$ term component bias and the magnitude of this bias-canceling effect. Bias canceling within Q^* is most apparent in cases in which $M(G^{\uparrow}_{LW})$ increases, yet $M(Q^*)$ decreases, as with the bulk scheme. Figure 9 shows the tendency for $M(Q^*;\text{bulk}) \lesssim M(Q^*;\text{PHX-A/B})$ for a given season and PBL-SLS, yet $M(Q^*;\text{PHX-A/B}) \lesssim M(Q^*;\text{bulk})$. However, $M(G^{\downarrow}_{LW})$, $M(G^{\downarrow}_{SW})$, and $M(G^{\uparrow}_{SW})$ are largely unchanged for a given season and PBL-SLS. Furthermore, Fig. 9 is useful for assessing where models perform poorly, by seeking variables and cases with large values of $M(\cdot)$. For example, $M(\langle Q_H \rangle) \geq 100\%$ for all cases.

b. *Effect of model spatial resolution*

Figure 9 also shows the influence of model domain resolution (Δ_H) on $M(\cdot)$. From these data, the effect of Δ_H on $M(\cdot)$ is insignificant, for all of the variables examined, aside from T_{2m} for summer YSU-MM5 PHX-A, where 333 m showed minor improvement over coarser resolutions (Fig. 9e). However, small differences

($<1\sigma$) in the mean values are observed between the different Δ_H of SEB variables for some cases (Fig. 9). From this insignificant convergence or divergence with varying Δ_H, we infer that surface-layer variables are not significantly sensitive to the aforementioned terra incognita resolution-dependent issues with turbulence closure at the temporal and spatial resolution of observations. We hypothesize that evaluation of models at finescales is limited when model resolution surpasses resolution of either the observation footprint or mixing length scales within physical parameterization schemes.

4. Conclusions

We have performed an evaluation of the WRF Model through a comparison of model simulations with observational data derived from a flux tower within a high-intensity residential LULC neighborhood within the Phoenix metropolitan area for a range of diurnal and seasonal solar forcing during calm weather periods. An analysis of diurnal and seasonal model errors was performed for T_{2m}, along with the radiative flux components and the SEB terms. Evaluation of WRF simulations with observations was performed for horizontal model grid resolutions Δ_H, ranging from 9 km to 333 m. We determined that it is computationally feasible to perform real-time simulations with 5 nested domains to $\Delta_H = 333$ m for 40 vertical levels using modern cluster architecture.

Diagnostic values such as T_{2m} are derived by an evaluation of the stability profile. Given the diurnal nature of surface-layer stability, and the fact that the profiles change their form with stability, bias may appear in one stability regime but not in another. Therefore, although a particular model may compare well with observed T_{2m} for a given stability regime, this evaluation is incomplete and possibly misleading without detailed SEB comparisons, as was demonstrated by the bulk cases. Furthermore, T_{2m} is influenced by bias canceling between the T_0, $\langle Q_H \rangle$, and C_H variables. For this reason, we show that solely evaluating model performance based upon diagnostic variables, such as T_{2m}, is not sufficient and can in fact lead to incorrect conclusions on model evaluation. We also show that model evaluation can benefit by enabling more detailed assessment of model errors when considering the individual downward, in addition to upward (Grimmond et al. 2011), shortwave and longwave radiation components, as opposed to just evaluating net radiation. Errors in Q^* may be hidden, since longwave and shortwave net radiation components may also give rise to bias canceling between individual components, with each term having different mechanisms of bias.

Little effect was produced with decreasing Δ_H in our analysis despite entering the terra incognita. This scale independence indicates that the modeled SEB terms are dominated by the local representation of the land surface and radiative forcing over any resolution-dependent turbulence dynamics influence. Furthermore, since the model does not change surface flux profile relationships at finer scales, any resolution-dependent dynamics that may be present within the simulation above the surface layer must have weakly coupled feedback to the surface layer. Detection of Δ_H sensitivity within our analysis is limited by the averaging time, and footprint of observations, which at 0.5–1 km is near the larger scale of the described terra incognita regime. Last, these results suggest that simulation at $\Delta_H = 333$ m seems to not clearly improve results with the parameterizations examined.

Our results indicate that YSU-MM5 tends to perform better than MYJ-eta, and that UCM performs better than bulk for SEB terms and $G^{\uparrow}_{LW}$. The UCM shows sensitivity to choice of f_{urb} value, which for some variables (e.g., $\langle Q_E \rangle$, $\langle Q_H \rangle$, ΔQ_s) had larger daytime influence than the PBL scheme. UCM (for PHX-A/B) and the bulk scheme, combined with YSU-MM5, give similar results for T_{2m}. The bulk scheme with MYJ-eta also performed well. However, when evaluating the diurnal cycle of other variables (e.g., $G^{\uparrow}_{SW}$, $G^{\uparrow}_{LW}$, u^*, $\langle Q_H \rangle$, $\langle Q_E \rangle$), it is clear that UCM performs better than bulk during daytime. This conclusion regarding bulk versus UCM cannot be drawn when only examining the mean diurnal error, suggesting that evaluating the diurnal cycle is needed for improved model assessment of SEB. Our results also indicate that evaluation or consideration of model configuration for arid cities needs to include SEB terms, not just T_{2m}. All model configurations should represent the urban heat island, since they all have urban representation. However, we did not analyze for this effect. Decreasing Δ_H below 1 km does not substantially improve simulation results with the PBL-SLS and urban parameterizations tested. This null result might be due to PBL scheme mixing and smoothing of small scales. Thus, a better parameterization adapted for subkilometer grid scales needs to be tested, and we suggest that, if improvement is sought at finer scales, parameterizations need to be adapted for these scales. However, further investigation of parameterization (such as employing turbulent kinetic energy closure) needs to be performed before any definitive conclusion regarding the benefit of finer-scale resolution on T_{2m} and SEB terms can be achieved.

Application of irrigation could provide starting points for improvement of $\langle Q_E \rangle$, $\langle Q_H \rangle$, and $G^{\uparrow}_{LW}$. It would be necessary to prescribe irrigation for both mesic residential classes and in agricultural areas. However,

availability of necessary gridded water-use input data may be problematic. Scenarios exploring the limiting case of maintaining or periodically recharging to maximum field capacity, or basing soil moisture upon vegetation wilting point, may provide alternative means to supplement available input water-use data. In particular, the impact from a water management perspective could be explored by imposing irrigation either with a constant daily input or with a seasonal daily input.

Evaluation of finescale modeling is ultimately limited by simultaneously controlling for 1) model physics parameterization and fundamental turbulence theory, 2) input data parameterization, 3) model resolution and filters, 4) observation resolution and siting, and 5) anthropogenic influence. The fifth limiting factor contains aspects of the first two factors. Salient anthropogenic factors include LULC modifications, particularly urbanization (buildings, impervious surfaces, modified landscapes, etc.) and cropland. Some specific aspects that are anticipated to play an important role in improving model predictions, and that are also in need of further investigation, are irrigation of croplands and vegetation; energy input and waste heat within the urban area; representation of urban parameters within models from values readily derived from observations; pollutants and air quality influencing radiative forcing and, to a smaller degree, air temperature. Model resolution was not seen to have a significant impact on SEB terms for the observation footprint considered.

Acknowledgments. This work was supported by the National Science Foundation (NSF) under EaSM Grants EF-1049251 awarded to Arizona State University (ASU) and EF-1049224 awarded to University of California, Riverside, and by NSF Grant DMS 1419593 awarded to ASU. Flux-tower data are available from Central Arizona–Phoenix Long-Term Ecological Research (site manager Phil Torrant) with funding provided by Grant CAP3: BCS-1026865 and by NSF via EaSM Grant EF-1049251. We also thank for their input on this work T. Volo, E. Vivoni, F. Salamanca, S. Earl, and the anonymous reviewers. We thank the WRF group at the National Center for Atmospheric Research for providing the WRF code.

REFERENCES

Arnfield, A. J., 2003: Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.,* **23,** 1–26, doi:10.1002/joc.859.

Barnosky, A. D., and Coauthors, 2012: Approaching a state shift in Earth's biosphere. *Nature,* **486,** 52–58, doi:10.1038/nature11018.

Burian, S. J., S. P. Velugubantla, and M. J. Brown, 2002: Morphological analyses using 3D building databases: Phoenix, Arizona. Los Alamos National Laboratory, Tech. Rep. LA-UR-02-6726, 65 pp.

Chen, F., and J. Dudhia, 2001: Coupling and advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.,* **129,** 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

——, and Y. Zhang, 2009: On the coupling strength between the land surface and the atmosphere: From viewpoint of surface exchange coefficients. *Geophys. Res. Lett.,* **36,** L10404, doi:10.1029/2009GL037980.

——, and Coauthors, 2011: The integrated WRF/urban modelling system: Development, evaluation, and applications to urban environmental problems. *Int. J. Climatol.,* **31,** 273–288, doi:10.1002/joc.2158.

Ching, J. K. S., 2013: A perspective on urban canopy layer modeling for weather, climate and air quality applications. *Urban Climate,* **3,** 13–39, doi:10.1016/j.uclim.2013.02.001.

Chow, W. T. L., D. Brennan, and A. J. Brazel, 2012: Urban heat island research in Phoenix, Arizona: Theoretical contributions and policy applications. *Bull. Amer. Meteor. Soc.,* **93,** 517–530, doi:10.1175/BAMS-D-11-00011.1.

——, T. J. Volo, E. R. Vivoni, G. D. Jenerette, and B. L. Ruddell, 2014: Seasonal dynamics of a suburban energy balance in Phoenix, Arizona. *Int. J. Climatol.,* **34,** 3863–3880, doi:10.1002/joc.3947.

Coutts, A. M., J. Beringer, and N. J. Tapper, 2007: Impact of increasing urban density on local climate: Spatial and temporal variations in the surface energy balance in Melbourne, Australia. *J. Appl. Meteor. Climatol.,* **46,** 477–493, doi:10.1175/JAM2462.1.

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.,* **46,** doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Fernando, H. J. S., 2008: Polimetrics: The quantitative study of urban systems (and its applications to atmospheric and hydro environments). *Environ. Fluid Mech.,* **8** (5–6), 397–409, doi:10.1007/s10652-008-9116-1.

——, 2010: Fluid dynamics of urban atmospheres in complex terrain. *Annu. Rev. Fluid Mech.,* **42,** 365–389, doi:10.1146/annurev-fluid-121108-145459.

Foken, T., 2008: The energy balance closure problem: An overview. *Ecol. Appl.,* **18,** 1351–1367, doi:10.1890/06-0922.1.

Fry, J., and Coauthors, 2011: Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogramm. Eng. Remote Sens.,* **77,** 858–864.

Georgescu, M., M. Moustaoui, A. Mahalov, and J. Dudhia, 2011: An alternative explanation of the semiarid urban area "oasis effect." *J. Geophys. Res.,* **116,** D24113, doi:10.1029/2011JD016720.

——, ——, ——, and ——, 2013: Summer-time climate impacts of projected megapolitan expansion in Arizona. *Nat. Climate Change,* **3,** 37–41, doi:10.1038/nclimate1656.

Grimmond, C. S. B., and Coauthors, 2010: The international urban energy balance models comparison project: First results from phase 1. *J. Appl. Meteor. Climatol.,* **49,** 1268–1292, doi:10.1175/2010JAMC2354.1.

——, and Coauthors, 2011: Initial results from phase 2 of the international urban energy balance model comparison. *Int. J. Climatol.,* **31,** 244–272, doi:10.1002/joc.2227.

Grossman-Clarke, S., J. A. Zehnder, W. L. Stefanov, Y. Liu, and M. A. Zoldak, 2005: Urban modifications in a mesoscale meteorological model and the effects on near-surface variables in an arid metropolitan region. *J. Appl. Meteor. Climatol.,* **44,** 1281–1297, doi:10.1175/JAM2286.1.

——, ——, T. Loridan, and C. Grimmond, 2010: Contribution of land use changes to near-surface air temperatures during recent summer extreme heat events in the Phoenix metropolitan area. *J. Appl. Meteor. Climatol.,* **49,** 1649–1664, doi:10.1175/2010JAMC2362.1.

Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.,* **132,** 103–120, doi:10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2.

——, Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.,* **134,** 2318–2341, doi:10.1175/MWR3199.1.

Hunt, J. C., Y. V. Timoshkina, S. I. Bohnenstengel, and S. Belcher, 2012: Implications of climate change for expanding cities world-wide. *Proc. Inst. Civil Eng. Urban Des. Plann.,* **166,** 241–254.

Janjić, Z., 2001: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso Model. NCEP Office Note 437, 61 pp.

Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.,* **43,** 170–181, doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.

Kusaka, H., and F. Kimura, 2004: Coupling a single-layer urban canopy model with a simple atmospheric model: Impact on urban heat island simulation for an idealized case. *J. Meteor. Soc. Japan,* **82,** 67–80, doi:10.2151/jmsj.82.67.

Liu, Y., F. Chen, T. Warner, and J. Basara, 2006: Verification of a mesoscale data-assimilation and forecasting system for the Oklahoma City area during the Joint Urban 2003 field project. *J. Appl. Meteor. Climatol.,* **45,** 912–929, doi:10.1175/JAM2383.1.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-*k* model for the longwave. *J. Geophys. Res.,* **102,** 16 663–16 682, doi:10.1029/97JD00237.

Moeng, C., J. Dudhia, J. Klemp, and P. Sullivan, 2007: Examining two-way grid nesting for large eddy simulation of the PBL using the WRF model. *Mon. Wea. Rev.,* **135,** 2295–2311, doi:10.1175/MWR3406.1.

Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the surface layer of the atmosphere. *Tr. Geofiz. Inst., Akad. Nauk SSSR,* **24** (151), 163–187.

Nordbo, A., L. Järvi, S. Haapanala, J. Moilanen, and T. Vesala, 2013: Intra-city variation in urban morphology and turbulence structure in Helsinki, Finland. *Bound.-Layer Meteor.,* **146,** 469–496, doi:10.1007/s10546-012-9773-y.

Oke, T., 1987: *Boundary Layer Climates.* Routledge, 435 pp.

Shin, H. H., and S.-Y. Hong, 2011: Intercomparison of planetary boundary-layer parametrizations in the WRF model for a single day from CASES-99. *Bound.-Layer Meteor.,* **139,** 261–281, doi:10.1007/s10546-010-9583-z.

Skamarock, W. C., and J. B. Klemp, 2008: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.,* **227,** 3465–3485, doi:10.1016/j.jcp.2007.01.037.

——, 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_bw.pdf.]

Smith, B. D., and M. A. Zeder, 2013: The onset of the Anthropocene. *Anthropocene,* **4,** 8–13, doi:10.1016/j.ancene.2013.05.001.

Vörösmarty, C. J., and Coauthors, 2010: Global threats to human water security and river biodiversity. *Nature,* **467,** 555–561, doi:10.1038/nature09440.

Wang, Z.-H., E. Bou-Zeid, S. K. Au, and J. A. Smith, 2011: Analyzing the sensitivity of WRF's single-layer urban canopy model to parameter uncertainty using advanced Monte Carlo simulation. *J. Appl. Meteor. Climatol.,* **50,** 1795–1814, doi:10.1175/2011JAMC2685.1.

Wyngaard, J., 2004: Toward numerical modeling in the ''terra incognita.'' *J. Atmos. Sci.,* **61,** 1816–1826, doi:10.1175/1520-0469(2004)061<1816:TNMITT>2.0.CO;2.

Xie, B., J. C. H. Fung, A. Chan, and A. Lau, 2012: Evaluation of nonlocal and local planetary boundary layer schemes in the WRF model. *J. Geophys. Res.,* **117,** D12103, doi:10.1029/2011JD017080.

Zhang, D., 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, doi:10.1175/1520-0450(1982)021<1594:AHRMOT>2.0.CO;2.