The International Urban Energy Balance Models Comparison Project: First Results from Phase 1

The International Urban Energy Balance Models Comparison Project: First Results from Phase 1

C. S. B. GRIMMOND,[a] M. BLACKETT,[a] M. J. BEST,[b] J. BARLOW,[c] J.-J. BAIK,[d] S. E. BELCHER,[c] S. I. BOHNENSTENGEL,[c] I. CALMET,[e] F. CHEN,[f] A. DANDOU,[g] K. FORTUNIAK,[h] M. L. GOUVEA,[a] R. HAMDI,[i] M. HENDRY,[b] T. KAWAI,[j] Y. KAWAMOTO,[k] H. KONDO,[l] E. S. KRAYENHOFF,[m] S.-H. LEE,[d] T. LORIDAN,[a] A. MARTILLI,[n] V. MASSON,[o] S. MIAO,[p] K. OLESON,[f] G. PIGEON,[o] A. PORSON,[b,c] Y.-H. RYU,[d] F. SALAMANCA,[n] L. SHASHUA-BAR,[q] G.-J. STEENEVELD,[r] M. TOMBROU,[g] J. VOOGT,[s] D. YOUNG,[a] AND N. ZHANG[t]

[a] *King's College London, London, United Kingdom*
[b] *Met Office, Exeter, United Kingdom*
[c] *University of Reading, Reading, United Kingdom*
[d] *Seoul National University, Seoul, South Korea*
[e] *Laboratoire de Mécanique des Fluides, CNRS-Ecole Centrale de Nantes, Nantes, France*
[f] *National Center for Atmospheric Research, Boulder, Colorado*
[g] *National and Kapodistrian University of Athens, Athens, Greece*
[h] *University of Łódź, Łódź, Poland*
[i] *Royal Meteorological Institute, Uccle, Belgium*
[j] *Ehime University, Matsuyama, Japan*
[k] *The University of Tokyo, Tokyo, Japan*
[l] *National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan*
[m] *University of British Columbia, Vancouver, British Columbia, Canada*
[n] *CIEMAT, Madrid, Spain*
[o] *CNRM-GAME Meteo France-CNRS, Toulouse, France*
[p] *IUM, CMA, Beijing, China*
[q] *Ben Gurion University of the Negev, Beer-Sheva, Israel*
[r] *Wageningen University, Wageningen, Netherlands*
[s] *University of Western Ontario, London, Ontario, Canada*
[t] *Nanjing University, Nanjing, China*

(Manuscript received 27 July 2009, in final form 4 February 2010)

ABSTRACT

A large number of urban surface energy balance models now exist with different assumptions about the important features of the surface and exchange processes that need to be incorporated. To date, no comparison of these models has been conducted; in contrast, models for natural surfaces have been compared extensively as part of the Project for Intercomparison of Land-surface Parameterization Schemes. Here, the methods and first results from an extensive international comparison of 33 models are presented. The aim of the comparison overall is to understand the complexity required to model energy and water exchanges in urban areas. The degree of complexity included in the models is outlined and impacts on model performance are discussed. During the comparison there have been significant developments in the models with resulting improvements in performance (root-mean-square error falling by up to two-thirds). Evaluation is based on a dataset containing net all-wave radiation, sensible heat, and latent heat flux observations for an industrial area in Vancouver, British Columbia, Canada. The aim of the comparison is twofold: to identify those modeling approaches that minimize the errors in the simulated fluxes of the urban energy balance and to determine the degree of model complexity required for accurate simulations. There is evidence that some classes of models perform better for individual fluxes but no model performs best or worst for all fluxes. In general, the simpler models perform as well as the more complex models based on all statistical measures. Generally the schemes have best overall capability to model net all-wave radiation and least capability to model latent heat flux.

Corresponding author address: Sue Grimmond, Environmental Monitoring and Modelling Group, Department of Geography, King's College London, London, WC2R 2LS, United Kingdom.
E-mail: sue.grimmond@kcl.ac.uk

DOI: 10.1175/2010JAMC2354.1

1. Introduction

The world's population has become increasingly urbanized: around 29% of the global population were urban dwellers in 1950, rising to 47% by 2000, and this proportion is predicted to rise to 69% by 2050 (UN 2009). Thus increasing numbers of people are impacted by weather and climate in urban areas. There is a growing requirement for accurate weather forecasts and climate change information within cities, and concurrent increases in computer capabilities allow greater spatial resolution within models. In combination, there is a greater proportion of the earth's surface being categorized as ''urban'' and there are a larger number of smaller grid boxes in atmospheric models in which urban areas need to be resolved.

The surface morphology (i.e., urban form) and presence of impervious building materials, sparseness of vegetation, and anthropogenic heat, water, and pollutant contributions each have a significant effect on the climate of urban regions, which lead to phenomena such as the urban heat island. Thus, effects of the urban surface on the fluxes of heat, moisture, and momentum need to be accounted for in the land-surface schemes used within numerical models, although the complexity of these schemes has to be balanced with their computational requirements. A fundamental aim of urban energy balance models is to accurately predict fluxes at the local scale (10^2–10^4 m). Some calculate additional terms including within-canyon air temperatures and wind speed, and facet surface temperature. A facet is a surface of the urban geometry that can be characterized by a single temperature and surface energy balance, and that can interact thermodynamically with other facets (e.g., a wall facet exchanging longwave radiation with the road facet; Fig. 1). The outputs from the model may be hourly or higher temporal resolution for the whole surface, or be facet/orientation-specific.

Models have been developed to incorporate urban features for different applications including global climate modeling (e.g., Oleson et al. 2008a,b); numerical weather prediction (e.g., Best 1998, 2005; Masson 2000; Chen et al. 2004; Harman and Belcher 2006; Liu et al. 2006); air quality forecasting (e.g., Martilli et al. 2003) and dispersion modeling (e.g., Hanna and Chang 1992, 1993); characterization of measurements (e.g., Krayenhoff and Voogt 2007); and water balance modeling (e.g., Grimmond et al. 1986; Grimmond and Oke 1991). Across these schemes a wide range of urban features are incorporated; the models have varying levels of complexity, and different fluxes modeled (Table 1; Figs. 1, 2).

In this paper, the methodology and initial results from the first international comparison of a broad range of urban land-surface schemes are presented. The requirements of a land-surface model from the perspective of an atmospheric model are considered; that is, surface fluxes of heat, moisture, and momentum. Thus, the fundamental requirement for the models to be included is that they simulate urban energy balance fluxes. The forcing data for the surface models are the same as that which would be provided by an atmospheric model; that is, the incoming shortwave ($K\downarrow$) and longwave fluxes ($L\downarrow$), air temperature, specific humidity, and the wind components. From these the outgoing radiative fluxes ($K\uparrow$, $L\uparrow$), net all-wave radiation (Q^*), turbulent sensible heat flux (Q_H), turbulent latent heat flux (Q_E), and net heat storage flux (ΔQ_S) are modeled. In this context, the net heat storage includes the energy storage within the buildings, the road and underlying soil, and for some models the air space within the street canyon (Grimmond and Oke 1999a). In the urban environment it is also useful to consider the anthropogenic heat flux (Q_F) in the surface energy balance (Oke 1988):

$$Q^* + Q_F = Q_H + Q_E + \Delta Q_S. \quad (1)$$

Features such as additional sources of energy (Q_F), presence of built and natural surfaces, the bluff body nature of the buildings, and existence of urban canyons, combine to change energy partitioning in urban areas. Thus significant modification to rural land parameterization schemes is needed. While many urban models have been evaluated against observational datasets (e.g., Grimmond and Oke 2002; Masson et al. 2002; Dupont and Mestayer 2006; Hamdi and Schayes 2007; Krayenhoff and Voogt 2007; Kawai et al. 2009), with some models even using the same observations, these comparisons have not been conducted in a controlled manner that allows robust model intercomparison. The objective here is to do just that: to undertake a staged and carefully controlled classification and comparison of urban energy balance models and their performance. An important objective also is to determine which approaches minimize the errors in the simulated fluxes.

2. The characteristics of urban energy balance models

Urban energy balance models can be classified in a number of ways (see also Grimmond et al. 2009a); for example, they vary in terms of the fluxes they calculate (''F'' in Figs. 1 and 2). While all the models examined here calculate $K\uparrow$, $L\uparrow$, Q^*, and Q_H, some do not model either Q_E or the Q_F, and some model neither. Here, a series of features are used to classify the approaches

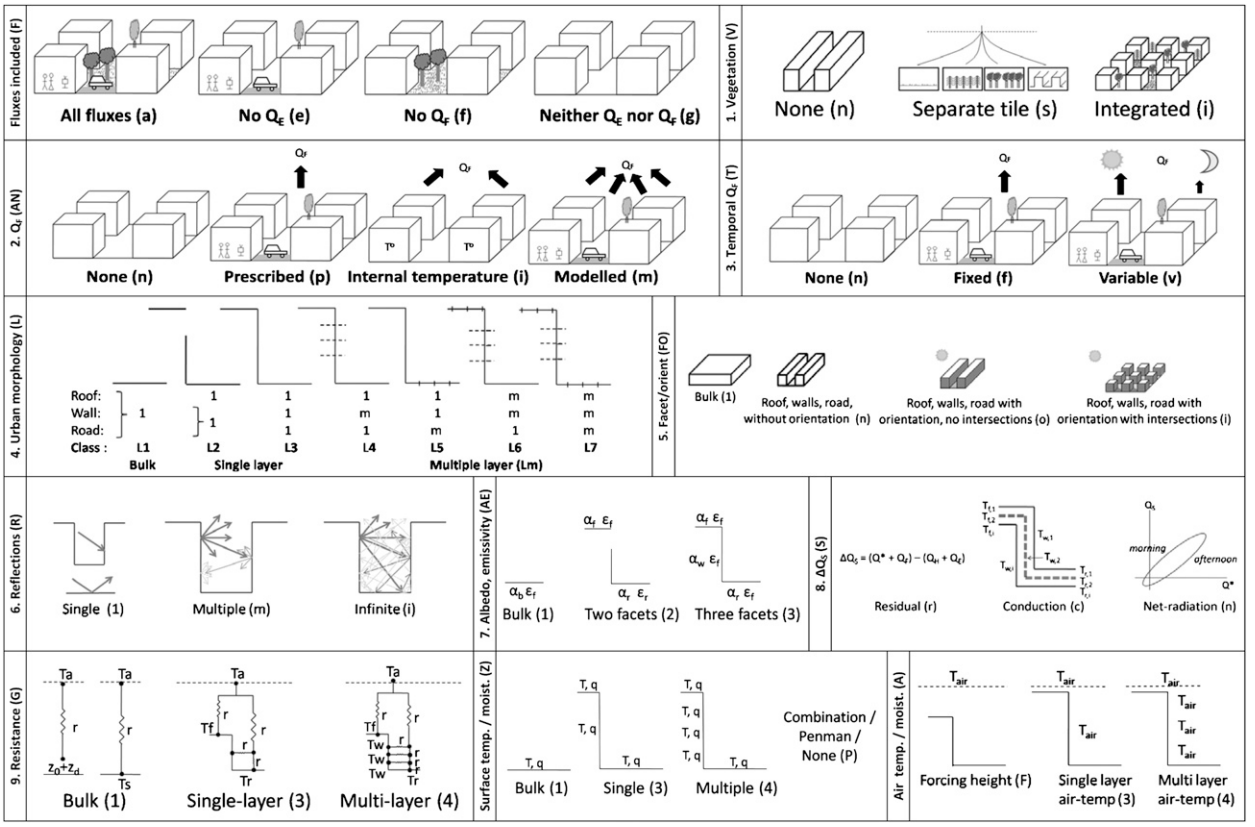

FIG. 1. Characteristics used to classify models (see Fig. 2).

taken. Figures 1 and 2 illustrate these, and the latter provides the numbers of models in each category. The illustrations also give each model class a reference in order to identify the category and its classification.

a. Vegetation and latent heat flux ("V" in Figs. 1 and 2)

A key decision in modeling an urban surface is whether or not vegetation (V) is simulated. A threefold classification is used here, where vegetation is

Vn: not considered,
Vs: modeled using a "tile" scheme to represent the surface heterogeneity (e.g., Essery et al. 2003) that does not interact with other surface types until the first atmospheric level of a mesoscale model (e.g., Best et al. 2006), and
Vi: "integrated" into the modeled urban surface.

The implication of not including vegetation is that there can be no latent heat except for periods immediately following rainfall. Some, even after rainfall, calculate no Q_E, whereas some account for dewfall and its later evaporation (Fig. 2b). For central business districts in many cities it may be reasonable to assume a negligible amount of vegetation and, hence, an absence of Q_E associated with vegetated surfaces. However, in residential areas (e.g., suburban North America) extensive fractions of the surface are vegetated, so the assumption of no urban Q_E is unrealistic. Moreover, in many locations, extensive street cleaning can result in water being available for evaporation despite the lack of vegetation (e.g., Mexico City, Mexico, Oke et al. 1999; Marseille, France, Grimmond et al. 2004).

The two classes of model that do incorporate vegetation differ in terms of the interactions which occur between the "built" and "vegetated" fractions (Figs. 2a–c). In the first case, "tiles" (Vs; Fig. 1), models typically take advantage of traditional land-surface schemes that have a wide variety of vegetation categories (e.g., Noilhan and Mahfouf 1996; Chen and Dudhia 2001; Essery and Clark 2003). Many have been extensively evaluated in the Project for Intercomparison of Land-Surface Parameterization Schemes (PILPS) (Henderson-Sellers et al. 1993, 2003; Irranejad et al. 2003) and other studies. Urban vegetation typically is more diverse than an individual vegetation class so a number of classes may be required (e.g., needleleaf and evergreen broadleaf trees)

TABLE 1. Urban energy balance models with the number of versions and number of groups utilizing each model participating in the comparison.

Code	Model Name	References	Versions	Groups
BEP02	Building Effect Parameterization	Martilli et al. (2002)	1	1
BEP_BEM08	BEP coupled with Building Energy Model	Martilli et al. (2002), Salamanca et al. (2009), Salamanca and Martilli (2009)	1	1
CLMU	Community Land Model–Urban	Oleson et al. (2008a,b)	1	1
GCTTC	Green Cluster Thermal Time Constant model	Shashua-Bar and Hoffman (2002, 2004)	1	1
IISUCM	Institute of Industrial Science Urban Canopy Model	Kawamoto and Ooka (2006, 2009a,b)	1	1
JULES	Joint U.K. Land Environment Simulator	Essery et al. (2003), Best (2005), Best et al. (2006)	4	2
LUMPS	Local-scale Urban Meteorological Parameterization Scheme	Grimmond and Oke (2002), Offerle et al. (2003)	2	1
NKUA	University of Athens Model	Dandou et al. (2005)	1	1
MORUSES	Met Office Reading Urban Surface Exchange Scheme	Harman et al. (2004a,b), Porson et al. (2009, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*)	2	1
MUCM	Multilayer Urban Canopy Model	Kondo and Liu (1998), Kondo et al. (2005)	1	1
NJU-UCM-S	Nanjing University-Urban Canopy Model-single layer	Masson et al. (2000), Kusaka et al. (2001)	1	1
NJUC-UM-M	Nanjing University-Urban Canopy Model-multiple layer	Kondo et al. (2005), Kanda et al. (2005a,b)	1	1
NSLUCM/NSLUCMK/ NSLUCM-WRF	Noah land surface model/Single-layer Urban Canopy Model	Kusaka et al. (2001), Chen et al. (2004)	3	3
SM2U	Soil Model for Submesoscales (Urbanized)	Dupont and Mestayer (2006), Dupont et al. (2006)	1	1
SNUUCM	Seoul National University Urban Canopy Model	Ryu et al. (2009)	1	1
SRUM2/SRUM4	Single Column Reading Urban Model tile version	Harman and Belcher (2006), Porson et al. (2009)	4	1
SUEB	Slab Urban Energy Balance Model	Fortuniak et al. (2004, 2005)	1	1
SUMM	Simple Urban Energy Balance Model for Mesoscale Simulation	Kanda et al. (2005a,b; 2007), Kawai et al. (2007, 2009)	1	1
TEB	Town Energy Balance	Masson (2000), Masson et al. (2002), Lemonsu et al. (2004)	1	1
TEB07	Town Energy Balance 7	Hamdi and Masson (2008)	1	1
TUF2D	Temperatures of Urban Facets 2D	Krayenhoff and Voogt (2007)	1	1
TUF3D	Temperatures of Urban Facets 3D	Krayenhoff and Voogt (2007)	1	1
VUCM	Vegetated Urban Canopy Model	Lee and Park (2008)	1	1

(a) Class details

Class code	Cap	VL92
Fluxes included (F)		
All fluxes (a)	14	10
No Q_E (e)	6	2
No Q_F (f)	13	16
Neither Q_E nor Q_F (g)	0	5
1 Vegetation (V)		
Not included (n)	9	14
Separate tile (s)	18	13
Integrated (i)	6	6
2 Q_F (AN)		
Negligible or ignored (n)	12	26
Prescribed (p)*	6	1
Internal Temp. (i)*	4	4
Modelled (m)*	5	1
i,p*	6	1
3 Temporal Q_F variation (T)		
None (n)	13	26
Fixed (f)	6	2
Variable (v)	14	5
4 Urban Morphology (L)		
Bulk (L1)	8	6
Single layer (L2)	5	7
(L3)	13	13
Multiple layer (Lm) (L4)*	2	2
(L5)*	1	1
(L6)*	1	1
(L7)*	3	3
5 Facets & orientation (FO)		
Bulk (1)	5	5
No orientation (n)	12	16
Orientation no intersec. (o)	10	6
Orientation & intersect. (i)	6	6
6 Reflection (R)		
Single (1)	11	11
Multiple (m)	15	15
Infinite (i)	7	7
7 Albedo, Emissivity (AE)		
Bulk (1)	5	6
Two facets (2)	6	4
Three or more facets (3)	22	23
8 ΔQ_S (S)		
Residual (r)*	6	7
Conduction (c)	24	23
Net radiation based (n)*	3	3

(b) Turbulent Flux Methods

	Q_H		Q_E	
	B	V	B	V
Resistance/Conductance (G)				
Bulk (1)	9	11	9	11
Single layer (3)	19	4	12	4
Multilayer (4)	5	2	3	2
Surface temp/moisture (Z)				
Bulk (1)	9	8	9	8
Single layer facet (3)	14	3	9	3
Multi layer facet (4)	8	4	4	3
Penman/comb no surface (P)	2	2	2	3
Air temperature/moisture (A)				
Single layer air (3)	24	14	17	14
Multi layer air (4)	6	2	4	2
Forcing height (F)	3	1	3	1
Other (O)				
Not included (N)	-	14	7	14
Specified/fixed evap. rate (E)	-	-	2†	2§
Soil only (W)	-	2	-	2
Human perspiration (H)	-	-	1	-

(c) Urban morphology (L) and vegetation (V) combinations (see b)

	B Q_H	V Q_H	B Q_E	V Q_E
L1	113 (4)	113 (2)	113 (2)	113 (2)
		N (2)	113 (2)	N (2)
	1P3 (2)	1P3 (2)	1P3 (2)	1P3 (2)
L2	113 (2)	113 (2)	113 (2)	113 (2)
	11F (1)	11F (1)	11F (1)	11F (1)
	31F (2)	N (2)	31F (2)	N (2)
	333 (2)	N (2)	N (2)	N (2)
L3	333 (9)	113 (3)	333 (3)	113 (3)
		133 (1)	333 (1)	133 (1)
		343 (2)	333 (1)	343 (1)
			343E (1)	3P3E (1)
		N (3)	333 (1)	N (1)
			N (2)	N (2)
	334 (1)	N (1)	334 (1)	N (1)
	343 (3)	343 (2)	343 (1)	343E (1)
			E (1)	343 (1)
		N (1)	343 (1)	N (1)
L4	444 (2)	W (2)	N (1)	W (1)
			H (1)	W (1)
L5	434 (1)	434 (1)	434 (1)	434 (1)
L6	444 (1)	N (1)	444 (1)	N (1)
L7	343 (2)	N (2)	N (2)	N (2)
	434 (1)	434 (1)	434 (1)	434 (1)

	B Q_H	V Q_H	B Q_E	V Q_E
Vn	113 (2)	N (2)	113 (2)	N (2)
	31F (2)	N (2)	31F (2)	N (2)
	333 (5)	N (5)	333 (1)	N (1)
			N (4)	N (4)
	334 (1)	N (1)	334 (1)	N (1)
	343 (3)	N (3)	343 (1)	N (1)
			N (2)	N (2)
	444 (1)	N (1)	444 (1)	N (1)
Vs	113 (4)	113 (4)	113 (4)	113 (4)
	11F (1)	11F (1)	11F (1)	11F (1)
	333 (4)	113 (1)	333 (1)	113 (1)
		133 (1)	333 (1)	133 (1)
		343 (2)	333 (1)	343 (1)
			343E (1)	3P3E (1)
	343 (1)	343 (1)	E (1)	343 (1)
	434 (1)	434 (1)	434 (1)	434 (1)
	444 (2)	W (2)	N (1)	W (1)
			H (1)	W (1)
Vi	1P3 (2)	1P3 (2)	1P3 (2)	1P3 (2)
	333 (2)	113 (2)	333 (2)	113 (2)
	343 (1)	343 (1)	343 (1)	343E (1)
	434 (1)	434 (1)	434 (1)	434 (1)

(d) Energy Balance Closure

		Yes	No		VL92 actual closure	
Forced		24	9		Yes	No
Closure forced by	Surface Temperature	6	Model checks closure			
	ΔQ_S residual	17	Yes	No	31	2
	Turbulent fluxes available energy limited	1	7	2		

FIG. 2. Characteristics used to classify the models (see Fig. 1). (a) Class code and number of models that have this capability (cap) and were applied this way for VL92. Classes with few models (asterisks) are amalgamated for analysis. (b) Approaches used to simulate the built (B) and vegetated (V) turbulent heat fluxes (Q_H, Q_E). The numbers are for the VL92 runs. (c) Combined features [from (b)] GZA or "Other" used in the turbulent flux modeling for the VL92 runs with the numbers in each class shown. (d) Energy balance closure approach. In (b), the dagger symbol indicates that one is also 343 and § indicates that one is 3P3 and one is 343.

to ensure adequate representation. In the tile approach, the built and vegetated fluxes are typically weighted by their respective plan area fractions to contribute to total fluxes (e.g., Lemonsu et al. 2004). The integrated case (Vi) is the most physically realistic as it allows for direct interaction of built and vegetated surfaces. This additional complexity may require increased computing resources and parameter values.

b. *Anthropogenic heat fluxes (''AN'' in Figs. 1 and 2)*

The magnitude of Q_F varies across a city. Typically it will be greatest in the densest part of the city (Oke 1988; Grimmond 1992; Ichinose et al. 1999). But even low absolute Q_F values may be important where they exceed the radiative forcing (e.g., cloudy, cold winters with low solar input).

Similar to Q_E, not all models consider Q_F. The four general approaches are

1) ANn: Q_F is assumed to be zero, negligible, or ignored;
2) ANp: Q_F is assumed to be a fixed amount that is required as specified input to the model, or is directly coded into the program;
3) ANi: Q_F is calculated based on assumed internal building temperature; and
4) ANm: Q_F is calculated and incorporates internal heat sources from buildings, and/or mobile sources associated with traffic, and/or metabolism.

Models that calculate Q_F typically include the heat related to internal heating of buildings as a minimum. A fixed temperature is assigned internally and this may, or may not, be allowed to vary seasonally or diurnally. Alternatively a fixed minimum (maximum) temperature is used so the internal temperature of the building may vary but within limits. The heat flux from traffic typically is based on assumptions about traffic flow from vehicle counts. The models that calculate Q_F in more detail, using a building energy model, mostly use the method of Kikegawa et al. (2003).

Beyond the internal temperature, the introduction of Q_F requires consideration of where the heat is released or added to the atmosphere; for example, whether heat is added within or above the canyon.

c. *Anthropogenic heat fluxes: Temporal variation (''T'' in Figs. 1 and 2)*

Anthropogenic heat flux Q_F varies both diurnally and seasonally (e.g., Sailor and Lu 2004; Offerle et al. 2005; Pigeon et al. 2007; Lee et al. 2009), although only some models consider this. Models that prescribe a fixed value (Tf) are likely to provide too much Q_F at night and insufficient quantities in the day; they will also not capture peak values normally associated with commuting (seasonally this peak may be associated with low solar radiation forcing). The inclusion of a diurnal and/or seasonal cycle (Tv) is more significant for certain applications when the modeled fluxes must be correct for specific short time periods. It is less significant when applications are not concerned with diurnal patterns.

d. *Urban morphology (''L'' in Figs. 1 and 2)*

Urban morphology affects radiative and turbulent heat exchanges. A number of approaches are used to capture these features, including

- L1: Slab or bulk surface;
- L2: single-layer approaches, which separate the surface into a roof and canyon (wall plus road) or
- L3: where the three facets (roof, wall, and road) are treated separately; and
- L4–L7: multiple-layer approaches, which divide one or more of the facets into layers or patches.

Slab models represent the urban form as a flat horizontal surface with appropriate ''bulk'' radiative, aerodynamic, and thermal characteristics. This has the advantage of simplicity and reduced computational time and parameter requirements. Single-layer models simplify the urban form to an urban canyon with a roof, wall, and a road. This allows for more realistic representations of radiative trapping and turbulent exchange (Masson 2000; Kusaka et al. 2001; Harman et al. 2004a,b; Lee and Park 2008). Parameter values are assigned for each facet and one set of energy exchanges per facet is modeled. Multilayer schemes divide the walls into a number of vertical and/or the roof and road into a number of horizontal patches; each with their own parameter values and energy exchanges modeled. For some models this allows for variable building height, and for others even differing roof, wall, and road characteristics. Note that the range of multilayer models L4–L7 is not exhaustive; rather it covers the range compared here.

e. *Urban morphology: Facets and orientation (''FO'' in Figs. 1 and 2)*

Models can be further subdivided by how urban canyon morphology, specifically the number of facets and orientations, are dealt with. Models include those that assume no facets (or orientation) and hence a bulk (or slab) surface (FO1), those that assume one infinitely long canyon (FOn), and those that have infinitely long canyons that run in two cardinal directions (FOo). The canyons may be fixed in orientation and neglect shading

or assume a random distribution of street canyons within the domain. Alternatively, the canyon may be modeled assuming two walls that have sunlit and shaded fractions that vary through the day and year. More realistic models also include an intersection between canyons (FOi), significantly increasing the number of the interactions with other facets that need to be computed.

f. Radiative fluxes: Reflections (''R'' in Figs. 1 and 2)

As $K\downarrow$ and $L\downarrow$ are provided, it is the $K\uparrow$ and $L\uparrow$ that are modeled. Beyond the morphology, and therefore the degree of detail needed for the surface parameters, the major differences relate to the number of reflections assumed—R1: single; Rm: multiple; Ri: infinite.

The single reflection model is the least computationally intensive and used in both slab and single-layer models. Models that simulate multiple reflections include both single-layer and multiple-layer models. Infinite reflections may be accounted for by slab, single-layer, and multilayer models.

For longwave radiation, slab models determine one surface temperature, whereas for facet-specific models, multiple surface temperatures are calculated (Figs. 2b,c). The surface temperatures then provide the forcing for Q_H and ΔQ_S.

g. Radiative fluxes: Albedo and emissivity (''AE'' in Figs. 1 and 2)

The albedo and emissivity values that determine the radiative fluxes may either be defined as a single value (bulk, AE1); as two facets, similarly to the L2 category (AE2); or may consist of combinations of various facets, analogous to the L3 (or greater) category of models (AEf).

h. Storage heat flux (ΔQ_S; ''S'' in Figs. 1 and 2)

The ΔQ_S is significant in urban areas given the materials and morphology of the urban surface (Grimmond and Oke 1999a). In urban models, it is determined in the following ways:

1) Sr: difference or residual of the energy balance,
2) Sc: solution of the heat conduction equation by dividing the facets into a number of thickness layers, and
3) Sn: function of Q^* and surface characteristics.

All three methods are used by slab or bulk models (Fig. 3). For all three, the ability to model the outgoing longwave radiation will impact the values obtained given the common need to model surface temperature.

For those models in which heat storage is calculated as the residual of the surface energy balance, assumptions as to which fluxes are included (specifically Q_F and Q_E) are important. The second method, the solution of the heat conduction equation, is used extensively by slab, single-, and multiple-layer models. It requires various parameters for each (sub) facet, including: number of layers, layer thickness, thermal conductivity, and volumetric heat capacity of the various layer materials (Table 2). The number of layers resolved varies between 1 and 48, and may be of fixed or variable thickness. Currently, none account for changing water content of built materials associated with rain, so the material parameters are static. Some solve the heat conduction equation using the force-restore method, while others solve the one-dimensional heat conduction equation.

The third approach is to use a fraction of Q^*(Sn). Some models take into account the diurnal pattern of the flux through the objective hysteresis model (Grimmond et al. 1991).

i. Other features

The following characteristics are not explicitly used to classify the models in this evaluation but are presented here because of differences between models. They do not necessarily result in the models being grouped differently to the classifications above; that is, models fall into some common groupings across model classes (Figs. 2, 3).

1) TURBULENT SENSIBLE (Q_H) AND LATENT HEAT (Q_E) FLUX

Typically surface resistance (or its inverse, conductance) schemes are used to model Q_H and Q_E (''G'' in Fig. 1, Fig. 2b). Depending on urban morphology, these consist of either single (G3) or multiple (G4) resistance networks, which account for the number of facets and layers that are resolved. Bulk models (G1) have the simplest resistance network (Figs. 1, 2b). A wide range of resistance schemes is used (e.g., Rowley et al. 1930; Clarke 1985; Zilitinkevich 1995; Guilloteau 1998; Harman et al. 2004b). To determine the resistance the wind profile within/above the canyon, roughness length, and displacement length or drag coefficients and atmospheric stability may be taken into account. Drag is either not considered or is calculated using roughness length, exponential wind profile, or distributed drag. Exchange between the canopy air and building surfaces may be parameterized by a roughness length approach or distributed sources of heat (generally in conjunction with a distributed drag approach).

The number of temperatures resolved, which drive the gradients, varies both for the surface and the air (within the canyon; ''Z'' and ''A'' respectively in Figs. 1 and 2b), and these are related to morphology and the number

	Vn	Vs	Vi	ANn	ANp	ANi	ANm	ANip	Tn	Tf	Tv	L1	L2	L3	L4	L5	L6	L7	FO1	FOn	FOo	FOi	R1	Rm	Ri	AE1	AE2	AEf	Sr	Sc	Sn	**Capability**
Vn				1		2	2	4	1	4	4	1	2	3			1	2	1		6	2	1	8		1	2	6		9		Vn
Vs				9	4	1	2	2	9	2	7	5	3	6	3			1	2	10	4	2	8	3	7	2	4	12	5	12	1	Vs
Vi				2	1	1	2		2		4	2		3		1			2	2		2	2	4		2		4	1	3	2	Vi
ANn	10	11	5						12			7	2	1	1			1	5	5		2	8	2	2	5	4	3	1	9	2	ANn
ANp	1										5		1	4						3	2		2	2	1			5	2	3		ANp
ANi	2	1	1							1	3				1	1		2			2	2		3	1			4	1	3		ANi
ANm		1									6	1		3	1		1			2	2	2	1	4	1			6	2	3	1	ANm
ANip	1									5	1		2	4						2	4			4	2		2	4		6		ANip
Tn	10	11	5	26								7	2	1	1			1	5	5		2	8	2	2	5	4	3	1	9	2	Tn
Tf	1	1			1	1							2	3	1					1	5			4	2		2	4	1	5		Tf
Tv	3	1	1			3	1	1				1	1	8	1	1	1	2		6	5	4	3	9	3			15	4	10	1	Tv
L1	2	2	2	5	1				5	1									5	3			6		2	5	2	1		5	3	L1
L2	4	3		7					7											3	2		2	2	1		4	1	1	4		L2
L3	5	5	3	12				1	12		1									6	5	1	3	7	2			12	2	10		L3
L4		2				1	1			1	1										2	1		1	2			3		2 1		L4
L5			1			1					1											1		1				1		1		L5
L6	1			1					1													1		1				1		1		L6
L7	2	1		1		2			1		2										1	2		3				3	1	2		L7
FO1	1	2	2	5					5			5											5			5				3	2	FO1
FOn	9	5	2	14	1			1	14	1	1	1	7	8									5	2	5		4	8	1	10	1	FOn
FOo	2	4		3		2	1		3	1	2			3	2			1					1	7	2		2	8	3	7		FOo
FOi	2	2	2	4		2			4		2			2		1	1	2						6				6	2	4		FOi
R1	2	7	3	10	1				10	1		6	2	3					5	5	1					5	2	4		8	3	R1
Rm	8	3	3	12		3			12		3		2	8		1	1	3		6	3	6					2	13	3	12		Rm
Ri	4	3		4		1	1	1	4	1	2		3	2	2					5	2						2	5	3	4		Ri
AE1	2	2	2	5	1				5	1		6							5	1			6							3	2	AE1
AE2	2	2		4					4				4							4			2	2						6		AE2
AEf	10	9	4	17		4	1	1	17	1	5		3	13	2	1	1	3		11	6	6	3	13	7				6	15	1	AEf
Sr		5	2	5		1	1		5	1	1		1	3	2			1		2	3	2		4	3			7				Sr
Sc	13	8	2	19		3		1	19		4	3	6	10		1	1	2	3	13	3	4	8	11	4	3	4	16				Sc
Sn	1		2	2	1				2	1		3							2	1			3			3						Sn
VL92 use	Vn	Vs	Vi	ANn	ANp	ANi	ANm	ANip	Tn	Tf	Tv	L1	L2	L3	L4	L5	L6	L7	FO1	FOn	FOo	FOi	Rs	Rm	Ri	AE1	AE2	AEf	Sr	Sc	Sn	

FIG. 3. Analysis of the individual characteristics (Fig. 2) gives this combination of approaches: (top right) the capability of the models and (bottom left) how the models were applied for the VL92 runs analyzed here. See Figs. 1 and 2 for explanation of class codes. For example, with respect to the capability of the 33 models: five that model vegetation in a separate tile (Vs), have slab morphology (L1); while in application to the VL92 dataset, only two were run this way.

TABLE 2. Urban parameters used by the number of models indicated for the VL92 runs and also those indicated as being applicable for the models. Each subscript refers to a separate parameter: f = roof; r = road; w = wall; v = pervious; t = tree; H = building; g = grass, s = soil, m = momentum; h = heat, u = urban; b = bulk. For some models § indicates state variables, a dagger indicates fundamental parameters, or an asterisk indicates derived parameters. Also noted is visible (VIS), ultraviolet (UV), and near infrared (NIR).

Name (units)	Parameter	No. models (VL92)	No. models (capability)
Radiative			
Albedo (-)	$\alpha_{f,r,w,v,t,g,b}$	24, 22, 19, 9, 6, 6, 4	24, 22, 19, 9, 6, 6, 4
VIS, UV, and NIR albedo of vegetation (-)	$\alpha_{\mathrm{VIS,UV,NIR}}$	1, 0, 3	2, 1, 4
Emissivity (-)	$\varepsilon_{f,r,w,v,t,g,b}$	23, 21, 19, 5, 6, 5, 2	23, 21, 19, 5, 6, 5, -
Roughness			
Roughness length above canyon (m)	z0m,c; z0h,c	10, 8	10, 10
Material roughness length for heat (m)	$\mathrm{z0h(mat)}_{f,r,w}$	14, 16, 6	4, 16, 6
Material roughness length for momentum (m)	$\mathrm{z0m(mat)}_{f,r,w}$	16, 14, 6	17, 14, 6
Effective roughness length for heat (m)	$\mathrm{z0h(eff)}_{f,r,w}$	—	5, 0, 2
Effective roughness length for momentum (m)	$\mathrm{z0m(eff)}_{r,w}$	—	5, 2
Roughness length of grass/tree surfaces (m)	$\mathrm{z0m}_{g,t}$	4, 5	—
Bulk roughness length (m)	z0b	1	2
Sublayer Stanton no. (-)	S_n	1	1
Zero-place displacement height (m)	$\mathrm{d0}_{f,r,w}$	2, 1, 1	2, 1, 1
Thermal characteristics			
Volumetric heat capacity (MJ K^{-1} m^{-3})	$c_{\mathrm{Pf},r,w,v,t,g,s}$	23, 24, 20, 8, 6, 5, 12	23, 24, 20, 8, 6, 5, 12
Thermal conductivity (MJ K^{-1} m^{-3})	$K_{f,r,w,v}$	22, 22, 21, 8	22, 22, 21, 8
No. of layers (-)	$\mathrm{nl}_{f,r,w,v,s}$	19, 18, 17, 3, 14	19, 18, 17, 3, 14
No. of walls modeled (-)	nw	17	17
Layer thickness (m)	$\mathrm{dl}_{f,r,w,v,s}$	22, 20, 2, 13	22, 20, 2, 16
Urban morphology			
Height of measurements/reference height (m)	Z_{ref}	15	16
Mean height (m)	$Z_{H}\dagger_{,g,t,\mathrm{all}}$	17, 4, 5, 5	17, 4, 6, 7
Avg building separation/canyon width (m)	$W_X\dagger$	8	8
Avg width of buildings (m)	$L_X\dagger$	8	8
Vertical floor density (building proportion > n stories) (-)	λ_B	3	3
Sky view factor (-)	SVF	1	1
Mean block length (m)	$L_Y\dagger$	3	3
Mean long axis azimuth of walls (m)	az	2	2
Canyon height-to-width ratio (-)	λ_s^*	2	11
Frontal area index (-)	λ_F^*	6	6
Wall to nonbuilt horizontal area (-)	λ_{wnb}^*	2	2
Plan area			
Fraction of area (-)	$\lambda_{f,r,v,g,t}^*$	19, 19, 10, 8, 6	19, 19, 10, 8, 6
Anthropogenic heat flux			
Total anthropogenic heat flux (W m^{-2})	$Q_{f,\mathrm{tot}}$	6	6
Sensible heat flux from vehicles (W m^{-2})	$Q_{H,\mathrm{traf}}$	4	4
Sensible heat flux from industry (W m^{-2})	$Q_{H,\mathrm{ind}}$	3	3
Temperature §			
Mean internal building temperature (K)	T_{int}	6	8
Deep temperature (K)	$T_{\mathrm{dr},s}$	6, 6	6, 8
Facet temperature (K)	$T_{f,r,w,s}$	6, 6, 5, 4	6, 6, 5, 6
Min; max internal building temperature (K)	$T_{\mathrm{int,min;max}}$	6, 3	6, 3
Vegetation/soil specific			
Vegetation wilting point (m^3 m^{-3})	S_{wilt}	9	10
Rooting depth of grass/tree (m)	$\mathrm{dr}_{g,t,v}$	3, 3, 2	4, 4, 3
Min stomatal resistance (s m^{-1})	R_{min}	5	5
Leaf area index of vegetation within the urban canyon (-)	LAI	6	7
Vegetation thermal inertia (J m^{-2} K^{-1} $s^{-1/2}$)	T_i	1	1
Parameters used in radiation stress function (-)	R_s	3	3
Parameter used in vapor pressure deficit function (-)	P_v	3	3
B parameter (-)	B	12	12
Sand/clay/loam/quartz content of soil (-)	$S_{s,c,l,q}$	6, 4, 3, 1	7, 7, 4, 2
Max vegetation canopy water holding capacity (mm)	$M_{t,g}$	6, 3	6, 4
Optimum temperature in temperature stress function (K)	T_{opt}	1	1
Coef for max interception water storage capacity (-)	I_s	1	2

TABLE 2. (*Continued*)

Name (units)	Parameter	No. models (VL92)	No. models (capability)
Ecosystem respiration parameter (-)	E_r	1	2
Ratio d(biomass)/d(LAI) (-)	d/d	1	2
e-folding time for senescence (-)	E_f	1	2
Cuticular conductance (-)	C_c	1	2
Max air saturation deficit (100 g kg^{-1})	A_s	1	2
Leaf area ratio sensitivity to nitrogen (m^2 kg^{-1})	S_N	1	2
Lethal min value of LA ratio (m^2 kg^{-1})	L_{la}	1	2
Nitrogen concentration of biomass (m^2 kg^{-1})	N_c	1	2
Root fraction (%)	R_f	1	2
Tree coverage (%)	T_c	1	1
Sunny spots (%)	S_s	1	1
Canopy solar absorptivity (-)	C_a	1	1
Canopy solar transmissivity (-)	C_m	1	1
Canopy thermal time constant (-)	C_t	1	1
Tree evaporative cooling coef (-)	E_c	1	1
Moisture availability			
Moisture availability (m^3 m^{-3})	$M_{f,r,s}$	3, 3, 7	4, 4, 8
Hydraulic conductivity of the soil at saturation (m s^{-1})	K_{sat}	13	13
Critical normalized soil water content for stress (m s^{-1})	S_{crit}	2	3
Air dry soil moisture content limit (m s^{-1})	S_{lim}	3	3
Soil suction experienced in the soil at saturation (m)	S_{suc}	9	10
Max soil moisture content (field capacity; m^3 m^{-3})	$M_{s,\text{max}}$	12	12

of vertical layers in the model. Many assume Monin–Obukhov similarity holds, which may not be applicable within the urban canyon or within the roughness sublayer (Roth 2000). However, given the lack of well-tested alternatives, currently, this may be the most appropriate approach.

As Q_H is calculated typically using surface temperature to force the gradient, a balance is inherent in the solution of surface temperature between the $L\uparrow$, Q_H, and heat conduction. Depending on the model objective, performance may be improved for one flux at the expense of another. Models that use a combination method (P, Penman–Monteith, or combination-type approach) do not need to determine the surface temperature to calculate Q_H, but still need to allow for the transport of heat away from the surface.

The approaches taken to model resistances (G), surface temperature (Z), and air temperature (A) result in a large number of combinations (Fig. 2c, expressed in GZA order). Here they are shown relative to the urban morphology classes (L1–L7; Fig. 1) and the vegetation class (Vn, Vs, Vi). The approach taken for each turbulent flux (Q_H, Q_E) for the built (B) and vegetated (V) part of the surface are shown. It is clear that the earlier classifications (Fig. 2a) do not produce common characteristics for these fluxes. Given the wide range of approaches, these are not investigated in further detail in this paper. Subsequent analysis of a larger dataset will investigate this. For the calculation of Q_H for the built (B) fraction of the surface, the two most common classes of the nine different combinations are

1) 333: single-layer resistance (G3), surface temperature (Z3), and air temperature (A3); and
2) 113: bulk resistance per facet (G1), surface temperature (Z1), and a single air temperature (A3).

For the vegetated surfaces the two most common classes for Q_H are N: Q_H is not calculated; and 113. For Q_E from built surfaces the predominant classes are N, 113, and 333; but also of note are those models that account for the evaporative loss of water in one time step immediately following precipitation with a fixed rate of evaporation (E). For Q_E from vegetated surfaces, the predominant classes are also N and 113. Two models, which do simulate Q_H and Q_E for vegetated area, account for evaporation from soil moisture only and not the loss of water through vegetation. In these cases the soil temperature and moisture profile are calculated using the approach of Tremback and Kessler (1985). In urban areas bare soil is rare, some sort of vegetation is most likely to be present.

2) ENERGY BALANCE CLOSURE

Not all models explicitly force or check that they have energy balance closure [i.e., that Eq. (1) holds; Fig. 2d]. Lack of closure may result from numerical instabilities or lack of precision in the code, from a lack of evaluation, or

from inconsistent assumptions. Closure may be forced in a number of ways: through the calculation of ΔQ_S at the end of each time step as a residual, by updating the surface temperature of the facets, or by restricting the turbulent heat fluxes to the available energy ($Q^* - \Delta Q_S$). Closure is an important issue when the land-surface scheme is part of a long-term climate model simulation; without it, there may be long-term bias in the model.

3) Anthropogenic water flux and other capabilities

Water can be added to the urban environment by human activity. Water is released by combustion processes, cooling towers, and by people, which is equivalent to the Q_F release (anthropogenic latent heat flux). One model takes into account the loss of water through perspiration (a source of Q_E). Given there are very few estimates of this term (Heiple and Sailor 2008; Moriwaki et al. 2008), and it is likely to be small in many settings, it is not surprising that it has not been included in most models. The term may be important in very dry areas (e.g., high-latitude cities in winter, hot dry cities) and in areas with excessive air conditioning. The second significant source of water comes via the pipe network, most typically as irrigation (e.g., garden sprinkling) or broken water pipes. In many suburban areas, if gardening is a common residential activity, this can be a large additional source of water relative to precipitation, especially during the summertime (e.g., Grimmond and Oke 1986; Grimmond et al. 1996). Estimating this component requires assumptions in the algorithms and/or the input data to define: 1) how much, and when, water is applied to the area and 2) where in the area it is released (e.g., to all vegetated surfaces or just to irrigated grass). The representation of this source is important (e.g., Mitchell et al. 2001) but has been considered in few models (e.g., Grimmond and Oke 1991).

The presence of snow cover will influence the energy balance of urban regions, affecting the albedo and, during periods of snowmelt, acting as a significant sink for latent energy (Lemonsu et al. 2008). For models with facets, the energy budgets of horizontal surfaces (roof and road) will be the most significantly affected, with additional energy budgets for these surfaces being necessary (Masson 2000).

j. Model uniqueness

Using the 31 individual characteristics to classify the models compared (Fig. 2a), 26 unique combinations occur (Fig. 3). This varies between model capability and actual use (demonstrated here for a dataset termed VL92; see section 4). For example, 21 models have the capability to account for Q_F but only 7 utilize this capability for the VL92 application. Although there are preferred approaches (e.g., for Q_F Tv over Tf), there is a notable diversity; models that have a similar approach for one aspect frequently use quite different approaches for other model components.

3. Model inputs

Inputs of three general types are required to model urban areas: 1) site parameters to describe the surface morphology and materials; 2) time series of atmospheric or forcing variables as boundary conditions; and 3) initial thermodynamic and moisture state conditions. The complexity of urban areas and diversity of surface description methods in the 33 models results in more than 145 (or >200 if individual layer values are considered) different parameters and state variables being needed for all of the models. The parameters fall into nine broad classes (Table 2). Some parameters, which are unique to individual models, can be derived from more basic parameters (Table 3; Fig. 4). Given the large effort needed to collect these data over the wide range of urban areas globally, or even within individual countries (e.g., Feddema et al. 2005; Ching et al. 2009), we encourage model developers to use common parameters. Also, it is important that the parameters are clearly defined and not open to misinterpretation (Loridan et al. 2010).

Morphometric parameters vary greatly, using either basic information (e.g., height, width) from which required parameters are calculated (e.g., canyon aspect ratio, sky view factor), or the ''higher'' level parameters as the inputs. Table 2 lists basic parameters from which higher-level parameters can be calculated.

Urban material related parameters are required to account for radiative transfer (e.g., albedo, emissivity) and thermal characteristics. Because of the different ways to describe the surface (Figs. 1, 2), there are varying numbers of models that use particular parameters (Table 2). All models use some form of albedo but this may be a single bulk albedo (α_b), or albedos for the roof (α_f), wall (α_w), road (α_e) and so on. Thermal properties are specified explicitly either relative to mass (specific heat capacity) or volume (volumetric heat capacity), or implicitly from model lookup tables.

As noted in Fig. 2, Q_F is dealt with in a variety of ways. For those using a fixed value, a model parameter has to be specified or alternatively, internal building temperature may need to be specified (in Table 2, it is included under temperatures but could be specified under Q_F).

Temperatures are required for many models. These may be initial state conditions (e.g., facet temperatures), which will evolve during the run, or require model spinup of sufficient time, or may be fixed for the duration of the run (e.g., deep soil temperature). In many applications, it

TABLE 3. Fundamental morphometric parameters (units of meters) that can be used to derive dimensionless morphometric parameters. Also shown are thermal parameters. Note many different names are used for the same parameters. Refer to Fig. 4 for further definitions.

Parameteralternative names	Symbol	Derivation
Fundamental morphometric parameters (m)		
Mean building height	z_H	
Mean building length Mean building width	$L_{\overline{XY}}$	$L_{\overline{XY}} = \frac{\sum_{i=1}^{n}(L_{Xi} + L_{Yi})}{2n} = \sqrt{L_{\overline{X}} L_{\overline{Y}}}$ n = all relevant directions for weighting
Mean canyon width Mean building separation	$W_{\overline{XY}}$	$W_{\overline{XY}} = \frac{\sum_{i=1}^{n}(W_{Xi} + W_{Yi})}{2n} = \sqrt{W_{\overline{X}} W_{\overline{Y}}}$
Mean block length Mean canyon length	$D_{\overline{XY}}$	$D_{\overline{XY}} = L_{\overline{XY}} + W_{\overline{XY}}$
Derived morphological parameters		
Canyon height-to-width ratio Canyon aspect ratio	λ_s	$\lambda_S = \frac{z_H}{W_{\overline{XY}}}$
Plan area ratio Roof area ratio Building coverage ratio Building fraction	λ_P	$\lambda_P = \frac{L_{\overline{XY}}^2}{D_{\overline{XY}}^2}$
Road area fraction	λ_r	$\lambda_r = \frac{W_X D_Y + W_Y L_X}{D_{\overline{XY}}^2}$
Frontal area index Building frontal density	λ_F	$\lambda_F = \frac{z_H L_{\overline{XY}}}{D_{\overline{XY}}^2}$
Wall to nonbuilt horizontal area	λ_{WNB}	$\lambda_{\mathrm{WNB}} = \frac{2(L_X + L_Y) z_H}{W_X D_Y + W_Y L_X}$
Thermal parameters: fundamental		
Density (kg m^{-3})	ρ	
Specific heat (J kg^{-1} K^{-1})	c	
Thermal conductivity (W m^{-1} K^{-1})	λ	
Thermal parameter: derived		
Volumetric heat capacity (J m^{-3} K^{-1})	$C = \rho c$	

is likely to be difficult to have realistic or observed values to meet the need for the temperature profile within a building or the soil to be prescribed. This may mean that some models require a long initialization period (spinup) to ensure that the temperature profiles are stable and representative of expected conditions.

For the models that use a vegetation tile, all the parameters required are not summarized in Table 2.

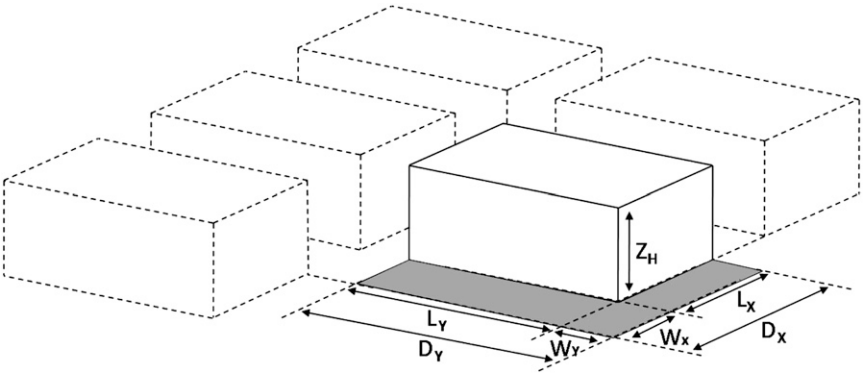

FIG. 4. Definitions of fundamental morphometric parameters (see Table 3). Note that with changing wind direction the frontal area index will vary.

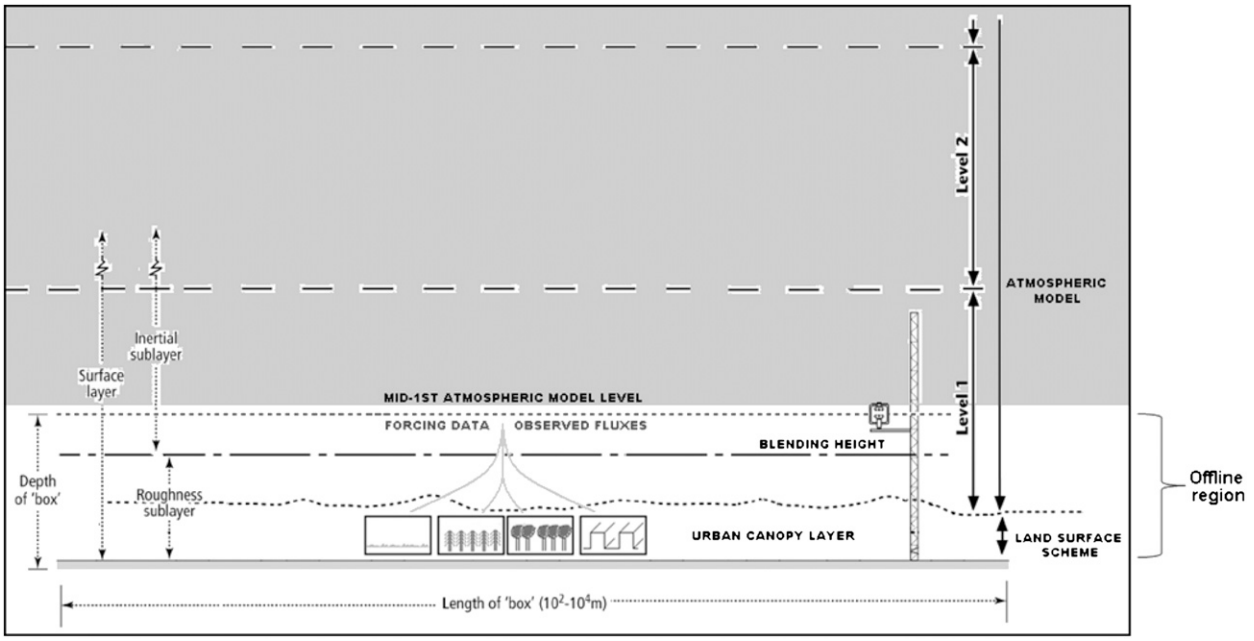

FIG. 5. Urban land-surface schemes simulate exchanges between the urban surface and the first layer of larger-scale atmospheric models. The observed fluxes and the forcing data are representative of the same level since they are above the roughness sublayer or blending height.

Parameter values, based on class selection, have been determined for extensive nonurban vegetated areas, and are assigned through model lookup tables. Model users have selected the vegetation class (e.g., grassland, deciduous or evergreen woodland, and/or bare soil) that they think is most appropriate in relation to the urban region they are modeling.

Soil moisture characteristics require both initial values and fixed parameters. These state variables have similar constraints and implications to that of the temperature. As urban areas often have disturbed soils and additional materials mixed into the media, it may mean that adoption of rural soil physical properties for parameters is not appropriate.

4. The International Urban Energy Balance Model Comparison Project

The methodology adopted here follows that of PILPS (Henderson-Sellers et al. 1993), which provided insight into both the models and real world processes. This allows the relative importance of key parameters to be determined and an assessment of the level of complexity required to produce reliable results. The International Urban Surface Energy Balance Model Comparison Project has been endorsed by the Global Energy and Water Cycle Experiment (GEWEX) Global Land–Atmosphere System Study (GLASS) and World Meteorological Organization Expert Team on Urban and Building Climatology.

The procedure for the comparison requires individual modeling groups (users and/or developers) to run their model(s) offline using forcing data provided for the top of the model, as would be provided by an atmospheric model (Fig. 5). This implies that parameter values should be representative of the observational footprint (see discussion in Masson et al. 2002). There is no feedback to larger-scale conditions within the modeling domain, so no larger-scale advection can occur, as would be present in a mesoscale or larger-scale model. The temporal resolution of analysis is typically 30 or 60 min, but individual models may be run at higher temporal resolution (1.5–300 s) and then average or sample their data back to the specified time interval of analysis (60 min). The spatial scale for both the measurements and models is the local or neighborhood scale (10^2–10^4 m). However there is no actual grid size because the models are run in single column mode. The observed fluxes and the forcing data are taken from tall towers that have the sensors located above the roughness sublayer (Grimmond and Oke 1999b; Roth 2000; Masson et al. 2002; Grimmond et al. 2004; Grimmond 2006). This height is equivalent to being above or at the blending height and is typically taken as the first atmospheric layer in mesoscale or larger-scale models (Fig. 5). The rationale for offline simulation is that although larger-scale circulation models may be accurate at the macroscale, their outputs will often be incompatible with those required as inputs to mesoscale urban surface models (Pitman et al. 1990). Equally, running such models

offline prevents feedbacks between climate and land surface, meaning that the sensitivity of the land-surface schemes themselves can be examined while the overlying atmospheric conditions are effectively held fixed (Wilson et al. 1987; Henderson-Sellers and Dickinson 1992).

The forcing data provided to participants were collected from a light industrial site in Vancouver, British Columbia, Canada (termed here VL92; Voogt and Grimmond 2000; Grimmond and Oke 2002). All observational data have measurement errors. These are associated with instrumental errors, instrument siting, fetch, flux corrections, lack of energy balance closure, and neglected terms etc. (e.g., Offerle et al. 2005; Grimmond 2006). This dataset was chosen as it has been used previously by a number of groups to evaluate their models (Grimmond and Oke 2002; Masson et al. 2002; Best et al. 2006; Krayenhoff and Voogt 2007; Oleson et al. 2008a). This meant that parameter values were reasonably well known. Also, the observed fluxes were provided so no model/group had an advantage from previous knowledge of this data.

The observations used in the evaluation consist of Q^*, Q_H, and Q_E plus ΔQ_S determined as a residual (Grimmond and Oke 1999a). During the observations (14 days in August 1992), the area was in drought and there was an irrigation ban in the city that was adhered to (Grimmond and Oke 1999c). The area is characterized by little vegetation (<5% plan area cover) and the soil moisture was very low at the time of data collection (Grimmond and Oke 1999a,c), making Q_E at this site small relative to the other fluxes (Table 4). The summertime conditions are expected to be associated with low Q_F as the area did not have extensive use of air conditioning or other significant sources of Q_F. This would be expected to be more significant in the winter but is not considered here as no observational data were available.

The purpose of this comparison is not to identify the best model, but to understand model errors related to the type of approach taken (Figs. 1, 2). Each model was assigned a random identifier number, which is used in the subsequent analysis of the results to ensure anonymity. The returned simulation data from each of these models were used to perform a series of statistical analyses to evaluate model performance (Table 5).

5. Results from VL92

Using the VL92 dataset, 33 different models/versions of models were analyzed (Table 1). Modeling groups assigned parameter values and initial state conditions they thought appropriate. Of the 33 participants, 20 chose to rerun their models subsequent to their initial submission and based on developments of their models during the period of the model comparison, thereby improving performance. Of those who did, 16, three, and one groups reran their models once, twice, and three times, respectively, with a decrease in the root-mean-square error (RMSE) in all cases except for the minimum values for Q_E and ΔQ_S, which remain the same (Table 4). The remainder of this paper evaluates the performance based on the final run results only.

As noted, this site has been used to evaluate model performance in previous studies (Table 4). These evaluations are not directly comparable to the current data as the same forcing data were not used in all the studies, and the time periods are not consistent, unlike the current comparison where all models have followed an identical protocol. However, comparing those results with the ''final'' runs presented here we can see that the results are similar. As with the overall cohort of models participating in the International Urban Model Comparison, there is some suggestion that model performance may have improved in the current (final) runs.

For Q^* the models, on average, have a smaller systematic RMSE (RMSE_S) than unsystematic RMSE (RMSE_U; Table 4). However, the maximum RMSE_S (81.9 W m^{-2}) is the same order of magnitude as the maximum RMSE_U (80.7 W m^{-2}), suggesting there are problems that could be fixed, for example, by changing parameter values. For Q_H the mean and maximum RMSE_S are larger than the RMSE_U, also suggesting that model results might be improved.

The ranked performance of the individual models, based on RMSE calculated for the 312-h dataset, for the four fluxes is shown in Fig. 6. No individual model performs best or poorest for all fluxes. For each flux, when models are ordered from best to poorer performance, in the better performing models there are small differences in RMSE. However, there is a point of step drop in performance: for Q^* five models performing less well; for Q_H, 15 models show distinctly poorer performance.

The encouraging performance for Q_E, with small RMSE values and only two models performing noticeably poorer, is a function of its small flux (Table 4). When a normalized Taylor (2001) plot is considered (Figs. 6e–h), where the ideal model would fall at the square (the observations), Q_E is the least well modeled (Fig. 6h). For Q^*, the models cluster most closely to the observed value, except for the five outliers already identified. Again for Q^*, all models have a correlation coefficient (r) of greater than 0.95 except for one, which has an r value of over 0.9. It is interesting that there is less of a step drop in ΔQ_S model performance but an almost constant correlation coefficient for all models (~0.9). Also Q_H has an almost constant correlation coefficient

TABLE 4. Summary of the mean, maximum, and minimum statistical performance (see Table 5 for definitions of statistics) across 33 models when compared with the VL92 dataset for all hours (n = 312 h). Also, RMSE statistics are displayed for the first run of the models. RMSE values from previous evaluations using VL92 data for all hours (note these are not directly comparable as different time periods and forcing data are used in some cases): 1 indicates Grimmond and Oke (2002); 2 indicates Masson et al. (2002) for periods 223–236, 225–231, and 232–236), respectively (the asterisk denotes combined $Q_H + Q_E$); 3 indicates Best et al. (2006) tile 1 and 2, respectively; 4 indicates Krayenhoff and Voogt (2007) LI1 (original simulation) and LI2 (with parameter adjustments), respectively, for TUF3D for 0300–2100 on day 227; 5 indicates Oleson et al. (2008b) for periods 225–231 and 232–236, respectively.

Statistic		Q^*	Q_H	ΔQ_S	Q_E
$\bar{x}_{obs}$ (W m^{-2})		131.2	71.8	42.4	15.5
σ_{obs} (W m^{-2})		217.6	89.2	133.0	21.0
$\bar{x}_{mod}$ (W m^{-2})	Max	193.6	208.5	83.9	30.7
	Min	84.4	49.8	−15.3	0
	Mean	133.7	113.6	13.3	7.4
σ_{mod} (W m^{-2})	Max	268.6	197.4	187.3	34
	Min	153.8	67.7	41.5	0
	Mean	231	120.5	119.2	8.8
R^2	Max	0.99	0.85	0.88	0.39
	Min	0.85	0.61	0.45	0.01
	Mean	0.98	0.8	0.79	0.25
RMSE prior runs (W m^{-2})	1	—	49	—	20
	2	59, 57, 59	76, 97, 50*	91, 105, 66	—
	3	69, 71	56, 43	103, 81	27, 24
	4	40.2, 31.1	138.5, 107.4	109.9, 98.1	—
	5	34, 34	81, 49	86, 59	16, 23
*RMSE first run (W m^{-2})	Max	177.9	233.3	311.4	157.4
	Min	28.4	39.3	49.1	17.2
	Mean	58.4	95.5	87.8	30.3
RMSE final run (W m^{-2})	Max	92.3	183.1	115.7	31.5
	Min	22.1	36.8	49.1	17.2
	Mean	47	81.7	77.8	23
RMSE$_S$ (W m^{-2})	Max	81.9	163.8	111.5	26.1
	Min	4.2	6.8	16	4.8
	Mean	30.3	58.9	54.8	19.8
RMSE$_U$ (W m^{-2})	Max	80.7	81.8	79.3	27.6
	Min	18.1	32.2	15.7	0
	Mean	33.6	53	50.6	7.4
MAE (W m^{-2})	Max	76.6	136.7	89.7	21.4
	Min	15.4	24.7	33.1	11.5
	Mean	37	57.1	57.5	15.6
MBE (W m^{-2})	Max	62.4	136.7	41.4	15.2
	Min	−46.9	−22	−57.7	−15.6
	Mean	2.7	44.1	−30.9	−8.3
d	Max	1	0.95	0.96	0.78
	Min	0.94	0.66	0.58	0.42
	Mean	0.99	0.86	0.88	0.54

for all models (~0.9). Based on the index of agreement d, on average model performance is best for Q^*, followed by ΔQ_S, Q_H, and Q_E (Table 4). This ranking is retained when the best overall performance (maximum d) of any model for each flux is considered.

Models need to respond to changes in exchange processes through the course of the day. Of interest, for example, is whether they resolve peak radiant and turbulent heat fluxes during the day as well as fluxes at night when shortwave radiation does not need to be considered. When the data are analyzed by time of day, RMSE is larger during the day (Fig. 7) as expected because of its larger absolute magnitude. Figure 7 shows results for three time periods: 1) day (1 h after $Q^* \geq 0$ W m^{-2}), 2) night (1 h after $Q^* \leq 0$ W m^{-2}), and 3) transition (remaining hours when Q^* is going through 0 W m^{-2}). The five models with the largest RMSE for daytime Q^* (Fig. 7), are the same as those for all hours (Fig. 6), although the ranked order differs slightly. The transition hours are particularly problematic for these models. The two poorest performing in the daytime are among the six-poorest performing at night.

The observed fluxes of Q_H may be underestimated on some occasions because of advection caused by sea

TABLE 5. Statistics used to analyze model performance (Willmott 1981; Jacobson 1999): P_i and O_i = predicted and observed values; n = number of data points; $e_i = P_i - O_i\hat{P}_i = a + bO_i$ (where a and b are the intercept and slope of regression line between O and P).

Statistic		Description/equation
$\overline{x}$	Mean	$\overline{P} = \dfrac{\sum_{i=1}^{n} P}{n}$
σ	Std dev	$\sigma_p = \sqrt{\dfrac{\sum_{i=1}^{n}(P_i - \overline{P})^2}{n-1}}$
R	Correlation coef (Pearson's)	$R = \dfrac{\sum_{i=1}^{n} O_i P_i - \dfrac{\sum_{i=1}^{n} O_i \sum_{i=1}^{n} P_i}{n}}{\sqrt{\left[\sum_{i=1}^{n} O_i^2 - \dfrac{\left(\sum_{i=1}^{n} O_i\right)^2}{n}\right]\left[\sum_{i=1}^{n} P_i^2 - \dfrac{\left(\sum_{i=1}^{n} P_i\right)^2}{n}\right]}}$
R^2	Coef of determination	R^2
RMSE	Root-mean-square error	$\mathrm{RMSE} = \left[n^{-1}\sum_{i=1}^{n}(e_i)^2\right]^{0.5}$
RMSE_S	Systematic RMSE	$\mathrm{RMSE}_S = \left[n^{-1}\sum_{i=1}^{n}(\hat{P}_i - O_i)^2\right]^{0.5}$
RMSE_U	Unsystematic RMSE	$\mathrm{RMSE}_U = \left[n^{-1}\sum_{i=1}^{n}(\hat{P}_i - P_i)^2\right]^{0.5}$
MAE	Mean absolute error	$\mathrm{MAE} = n^{-1}\sum_{i=1}^{n}\lvert e_i\rvert$
MBE	Mean bias error	$\mathrm{MBE} = n^{-1}\sum_{i=1}^{n} e_i = \sum\overline{P} - \overline{O}$
d	Index of agreement	$d = 1 - \dfrac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(\lvert O_i - \overline{O}\rvert + \lvert P_i - \overline{O}\rvert)^2}$

breezes (Masson et al. 2002). For Q_H the daytime errors are largest. At night the models generally do well almost across the board but the absolute values of the fluxes are smaller (Fig. 7). The daytime RMSE for Q_H is larger than for Q^* for all models. The $\mathrm{RMSE_S}$ tends to be greater for Q_H than for Q^*. For the most poorly performing models, $\mathrm{RMSE_s}$ is generally larger than $\mathrm{RMSE_U}$ (Fig. 7, circles plot above triangles).

Using the model classifications (Figs. 1, 2) we can evaluate whether particular approaches result in clear improvements in performance. It should be noted that the options used by groups were not always their most complex (cf. capability with VL92 options used in Figs. 2, 3). Two sets of statistics are used: RMSE and the mean bias error (MBE) for day and night (Figs. 8, 9) with results for each model shown as a point for each class and category (Fig. 2). The range, interquartile range (IQR), and mean and median performance of the category within the class can be compared. Perfect performance would have an RMSE and MBE of 0 W $\mathrm{m^{-2}}$. Given the relative magnitude of the MBE for nighttime Q_E ($<|12|$ W $\mathrm{m^{-2}}$), these results are not considered further here.

First, the method to represent vegetation (V class 1) is considered. Of the 18 models that have the ability to include vegetation as a separate tile (Vs; Fig. 2), five did not. Six additional models have integrated vegetation (Vi) within their urban surface. For the VL92 runs, a total of 14 models do not consider vegetation (Vn). The IQR of RMSE (bars on Fig. 8) is smaller in the daytime for Q^*, Q_H, and ΔQ_S when vegetation is included as a separate tile (Vs). In the daytime, not including vegetation (Vn) results in the largest RMSE medians (Q_H = 181, ΔQ_S = 136, Q^* = 59, and Q_E = 36 W $\mathrm{m^{-2}}$) and MBE medians (Q_H = 158 W $\mathrm{m^{-2}}$, $\Delta Q_S = -107$, $Q^* = 42$, and $Q_E = -28$). For daytime Q_H and Q^*, the tiled approach (Vs) has the smallest RMSE (median = 71 and 46 W $\mathrm{m^{-2}}$, respectively) and MBE (median = 18 and -14 W $\mathrm{m^{-2}}$, respectively), whereas the integrated vegetation (Vi) has the lowest individual RMSE values for Q^* and ΔQ_S. For daytime Q_E, the RMSE and MBE are best for Vi (median = 27 and 3 W $\mathrm{m^{-2}}$, respectively). At night for Q_H, the performance is poorest for those models that assume a separate tile (Vs; median RMSE = 19 W $\mathrm{m^{-2}}$, MBE = 17 W $\mathrm{m^{-2}}$) and best for Vi models

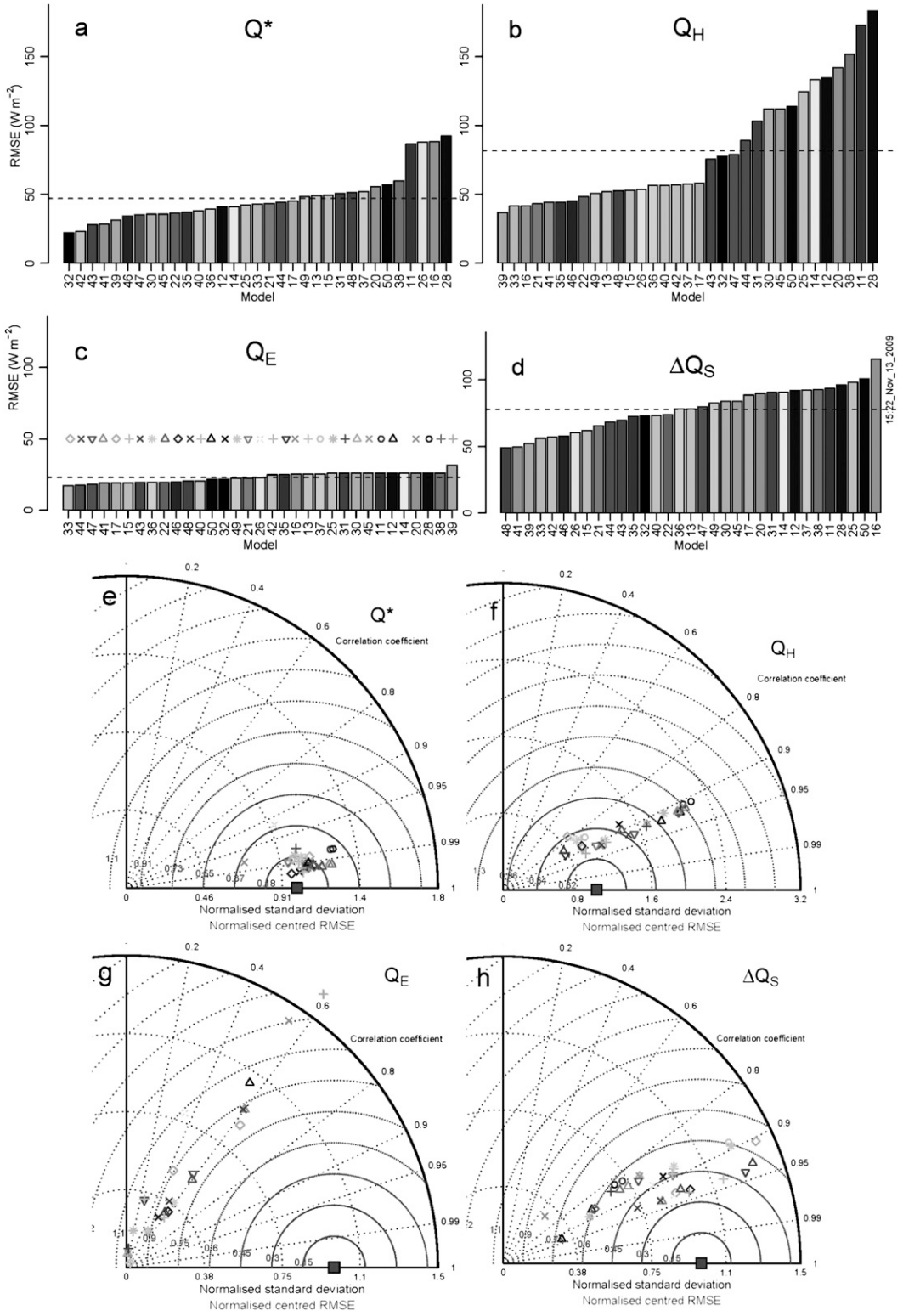

FIG. 6. (top) Ranked RMSE (W m^{-2}) and (bottom) normalized Taylor diagrams associated with each model for the whole time period. Each model is randomly assigned a number and symbol. The key for the symbols is shown in (c). (a),(e) Net all-wave radiation, (b),(f) turbulent sensible heat, (c),(g) latent heat, and (d),(h) storage heat fluxes. The dotted line is the mean RMSE. The Taylor plots display the correlation coefficient in relation to the polar axis comparing hourly values, the normalized standard deviation in relation to the horizontal axis and the normalized RMSE in relation to the internal circular axes (Taylor 2001); (N = 312 h).

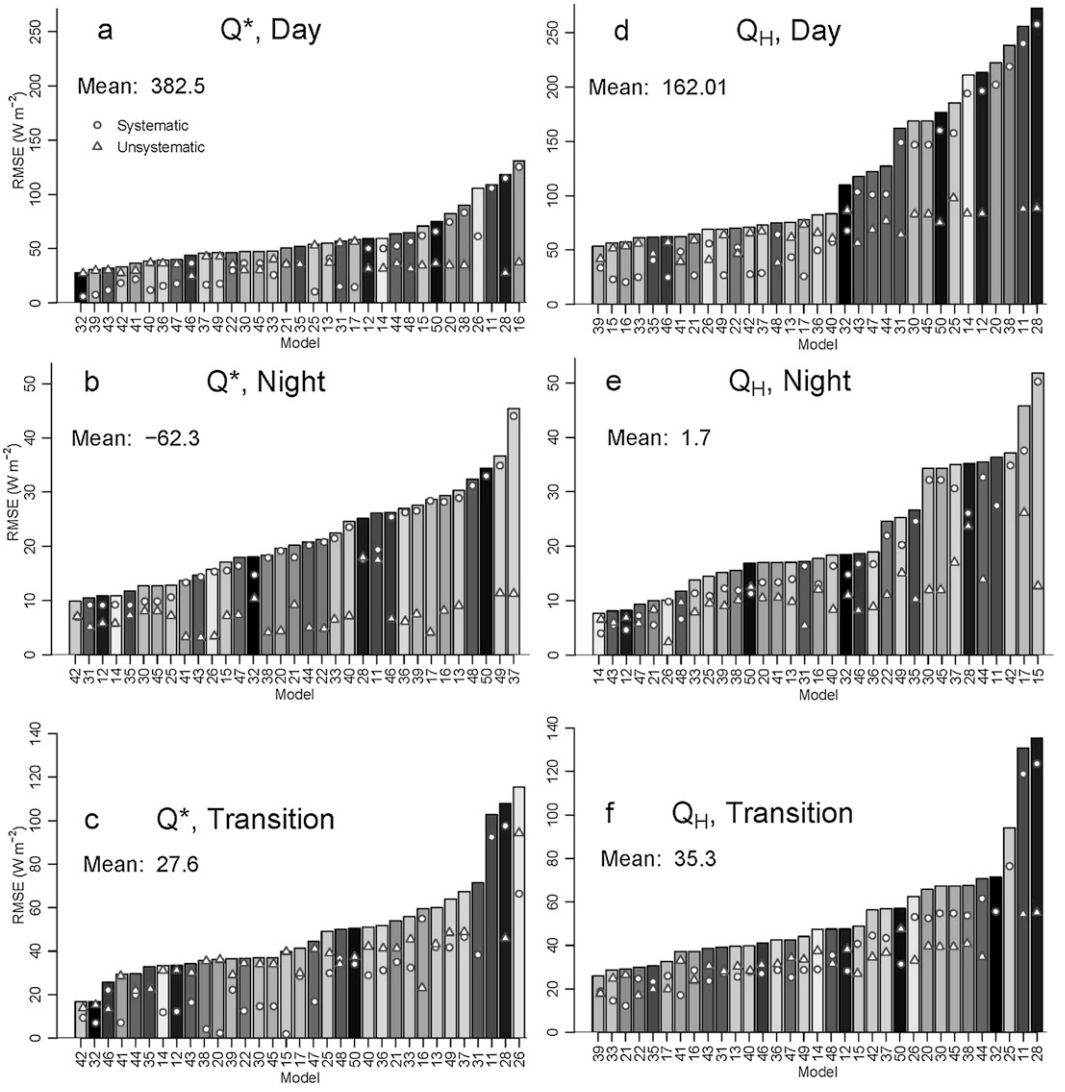

FIG. 7. Ranked RMSE (W m^{-2}) for (left) net all-wave radiation and (right) turbulent sensible heat flux by time of day (see text) for (a),(d) day, (b),(e) night, and (c),(f) transition time periods. Circles and triangles are $RMSE_S$ and $RMSE_U$, respectively. The mean observed flux (W m^{-2}) is for each period is given.

(median RMSE = 14 W m^{-2}, MBE = −2 W m^{-2}). However, for Q^*, Vi and Vn have similar performance (median RMSE Vi = 16, Vn = 19; MBE: Vi = −11, Vn = −8 W m^{-2}).

Examining the combination of model characteristics (Fig. 3) shows that for those that do not take into account vegetation, Vn, share only one common characteristic: their calculation of ΔQ_S via conduction or net radiation (class 8, Sc, Sn). However, many models that do include vegetation (Vs) also use this approach to heat conduction (Sc), so this is not likely to be a primary coexplanation. Not including vegetation even in this area where there is very little, and where the measured Q_E is small relative to the other fluxes, appears to impact the ability to model Q^* and Q_H, with a resulting poor performance also in ΔQ_S.

The VL92 site also has low Q_F. Most groups assumed it is negligible (ANn) with only seven groups explicitly including the flux (Fig. 2). Those that have considered it have taken a wide range of approaches but because of the small numbers they are grouped together into one class for analysis (ANm). Similarly, different temporal approaches to modeling Q_F (Tf, Tv) are used but the small number of models per class means analysis is the same and so is not shown separately. In the daytime, median RMSE and MBE are smallest for all fluxes when Q_F is ignored (ANn). This differs for nighttime fluxes however, where ANm models have

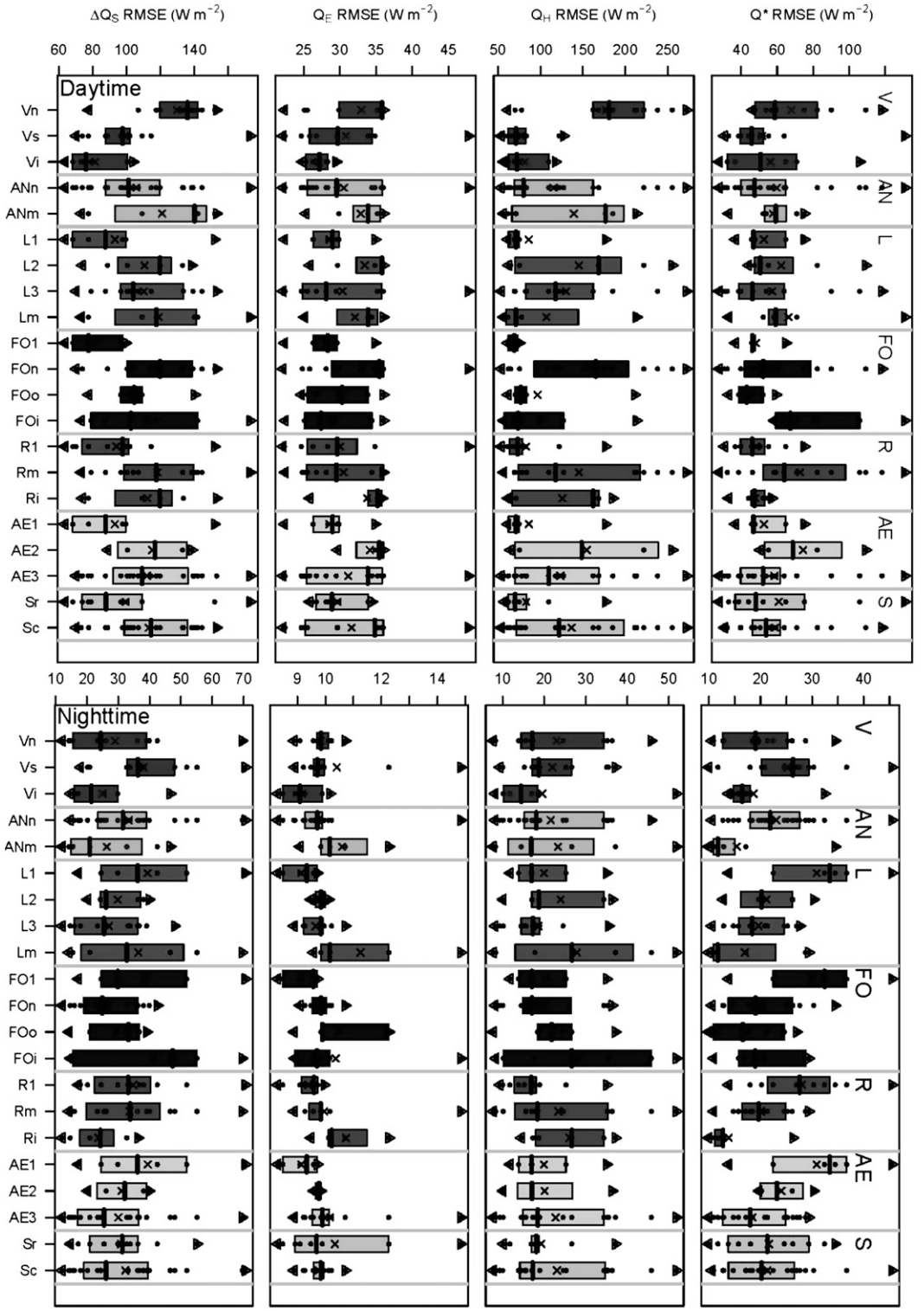

FIG. 8. RMSE (W m^{-2}) for each of the seven categories (see Fig. 2 for key) (left to right) ΔQ_S, Q_E, Q_H, and Q^* for (top) day and (bottom) night. Each dot is a model, the shaded bar shows the 25th and 75th percentile, and the line indicates the median and the "x" the mean. The maximum and minimum are indicated by the triangles. Note that the vertical scale varies between graphs. Mean observed fluxes for Q_E is 28.6 (day) and 3.9 W m^{-2} (night) and for ΔQ_S is 191.7 (day) and −70.7 W m^{-2} (night). For Q^* and Q_H see Fig. 7.

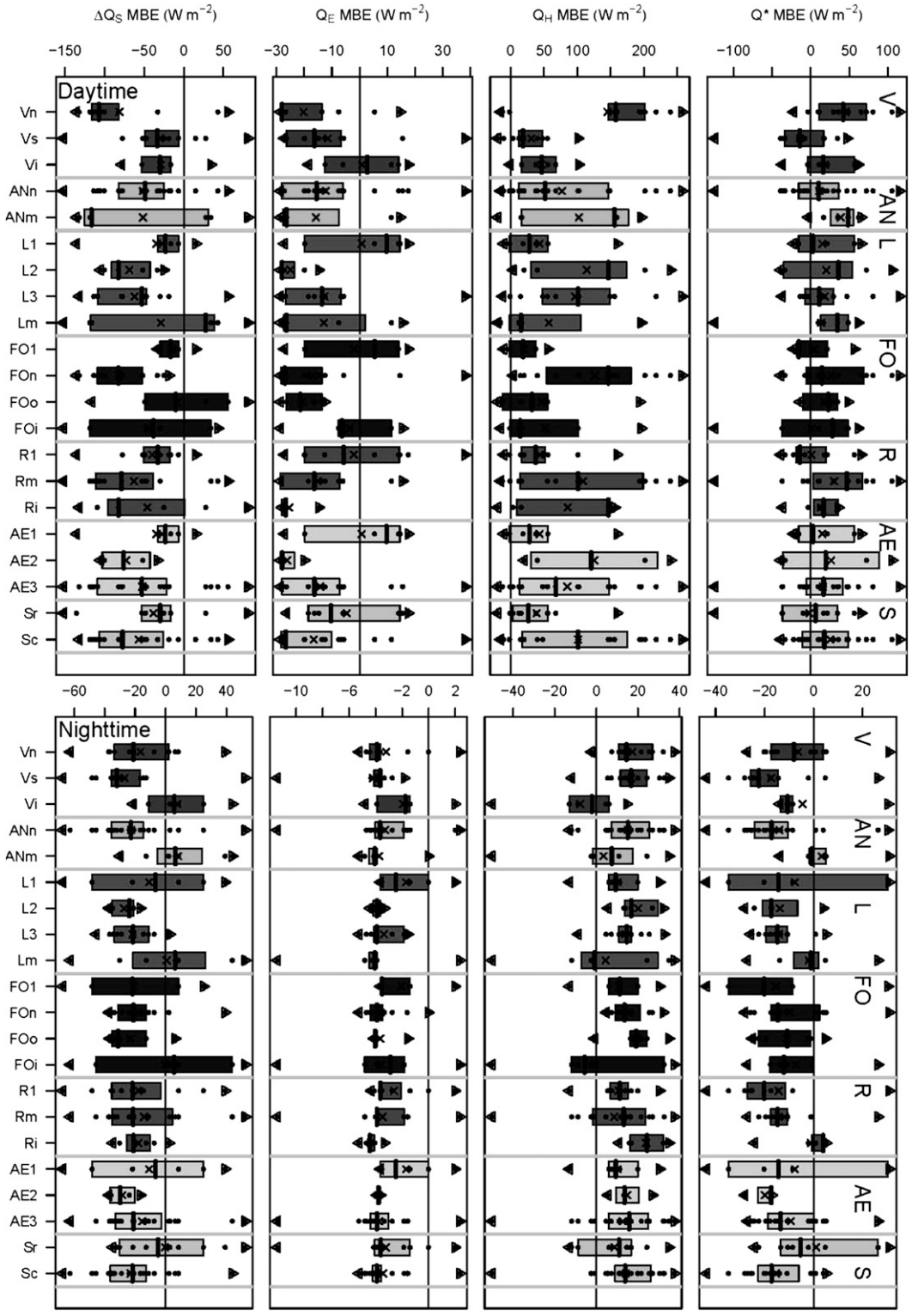

FIG. 9. As in Fig. 8 but for the MBE (W m^{-2}).

the smallest RMSE and absolute MBE for all fluxes except Q_E.

The model combinations in Fig. 3 show that those models that use an internal temperature (ANi) tend to have a fixed or variable temporal variation in Q_F (class 3, Tf, Tv), an urban morphology that is multilayered (L4–L7), and a surface albedo–emissivity that has three or more facets (class 7, AEf).

The urban morphology (class 4, L) has a relatively large within-class difference (range of median RMSE: 98 W m^{-2} and MBE: 130 W m^{-2}) for the daytime Q_H. For both RMSE and MBE, there is no clear best performer of models across all fluxes (best for RMSE median: Q^* = L3, Q_H = Lm and L1, Q_E = L3, ΔQ_S = L1; best for MBE median: ΔQ_S = L1, Q_H = Lm, Q^* = L1, Q_E = L1; Figs. 8, 9). At night, multilayer models (Lm = L4–L7) perform best for Q^*, Q_H, and ΔQ_S based on MBE (median $Q^* = -1$, $Q_H = -1$, ΔQ_S = 6 W m^{-2}). The urban morphology classes have few common characteristics, although all L1 models use a single reflection and a bulk albedo and emissivity. In addition, and by definition, L3 and all Lm models have three facets for albedo and emissivity.

With respect to the categorization based on facets and orientation (FO class 5), the largest difference is for the simulation of daytime Q_H (difference between category medians ΔRMSE of 96 W m^{-2}, ΔMBE = 129 W m^{-2}). Those that treat the surface as a whole (FO1) have the lowest daytime RMSE for Q_H and ΔQ_S (although for Q_H, median RMSE for FO1, FOo, and FOi differ by $<$8 W m^{-2} while it is lowest for FOo and Q^* and for FOi and Q_E). At night, the lowest median RMSE is: Q^* = FOo, Q_H = FO1, and FOn, Q_E = all groups equal, ΔQ_S = FOn. There is no consistency in groupings with the smallest daytime MBEs (Q^* = FOn, Q_H = FOi, Q_E = FO1, ΔQ_S = FOo). Except for Q^*, during the daytime, models that simulate a canyon but have no associated orientation (FOn), have the largest biases (Q_H: positive bias, Q_E and ΔQ_S: negative bias) and these are likely to be complementary. At night, models that incorporate orientation and intersections (FOi) have the smallest bias, again except for Q^*, where it is FOo models (although differing by just 1 W m^{-2} when compared with FOi). In the daytime, for Q_H, the median RMSE performance improves from FOn, FOo, FOi, and FO1 (165, 77, 74, and 69 W m^{-2}, respectively); and for Q^*, improves from FOi, FOn, FO1, and FOo (67, 52, 46, and 43 W m^{-2}). The unique combinations that these categories of models have in common include those that treat the surface as a whole (FO1), have no anthropogenic heat fluxes calculated (ANn) and, obviously, have just a slab surface morphology, single reflections, and a bulk albedo and emissivity. Models that include orientation (FOo, FOi) all assume three or more facets for albedo and emissivity (AEf; as would be expected). Models without orientation (FOn) largely utilize conduction methods to calculate the storage heat flux (Sc).

When the models are classified based on the number of reflections used, there are large within class differences (ΔRMSE = 89 W m^{-2} for daytime Q_H; Fig. 8). This is also the largest difference for the MBE (ΔMBE 109 W m^{-2}; Fig. 9). During the day, models with a single reflection scheme (class 6, R1) perform best for all fluxes except Q_E (median RMSE ΔQ_S = 98, Q_H = 73, Q^* = 46 W m^{-2}). The daytime MBE is smallest for Q^* models that calculate single reflections (Rs; median MBE Q^* = −14). Generally, during the daytime the models that have infinite reflections (Ri) perform least well for Q_H and Q_E, (median MBE Q_H = 147, $Q_E = -27$ W m^{-2}; median RMSE Q_H = 162, Q_E = 35 W m^{-2}); there are also negative median MBEs for all classes for Q_E and ΔQ_s, while Q_H and Q^* have a positive bias, with the exception of the single reflection class and Q^*. This suggests that the single reflection models may not allow enough radiation to be absorbed in comparison with observations. For ΔQ_S and Q_H, RMSE increases with the number of reflections modeled.

At night, models using increasing numbers of reflections have smaller RMSE for Q^* (Q^*: Ri = 13, Rm = 20, R1 = 28 W m^{-2}); whereas, the situation reverses for Q_H and Q_E, with those modeling fewer reflections yielding better results (Q_H: Ri = 27, Rm = 18, R1 = 17 W m^{-2}). For the calculation of Q^* at night, the Ri type models perform best with the lowest median RMSE and MBE (RMSE = 13, MBE = 4 W m^{-2}). However, as for daytime, superior performance for one flux is accompanied by poorer performance in another. All approaches have a similar sized negative MBE for nocturnal ΔQ_S (median from −21 to −22 W m^{-2}). The MBE for single reflections suggests that the surface temperature is too high, but correcting the bias during the daytime is likely to increase the nocturnal surface temperature, so there may be other issues with the models that use this method. Compensation also occurs between Q^* and Q_H most particularly at night. All schemes with infinite reflections (Ri) have three facets for albedo and emissivity (AEf).

The differences within groups of models are amongst the greatest when stratified based on specification of albedo/emissivity (class 7, AE). In general, using a bulk albedo–emissivity (AE1) results in better performance for all fluxes during the day based on median RMSE and MBE (median MBE $\Delta Q_S = -23$, $Q^* = 3$, $Q_E = 10$, Q_H = 28 W m^{-2}). Models using two facets (AE2) tend to have the poorest daytime performance (except for Q^* where median MBE for all groups is similar). At night, the differences in median MBE are smaller (Q_H:

AE2 = 14, AEf = 16, AE1 = 9; Q^*: AE1 = −14, AEf = −13, AE2 = −17 W m^{-2}). In this evaluation, where buildings are small and widely spaced, the ability to distinguish different facet characteristics of albedo and emissivity is not important. However, where buildings are taller, more tightly spaced, and/or constructed with very different materials, this result may not necessarily be the same. It is also important to remember that depending on the intended application, the ability to change facet material characteristics may be very important; for example, for scenario testing (e.g., for urban heat island mitigation).

Classifying models based on method used to calculate ΔQ_S (S class 8) has a relatively small difference in the median RMSE and MBE for all fluxes. Again the biggest difference in performance is associated with daytime Q_H (52 and 74 W m^{-2} for RMSE and MBE). The daytime Q^* differences are 6 and 12 W m^{-2} for RMSE and MBE, respectively; these are the smallest within-group differences in median for Q^* across the classes. The residual method (Sr) performs better during the daytime for all fluxes except Q^* daytime (median MBE: Q_H = 27, Q^* = 6, Q_E = −11, ΔQ_S = −30 W m^{-2}) and for all nighttime fluxes (median MBE: Q_H = 11, Q_E = 4, ΔQ_S = −5, Q^* = −5 W m^{-2}). The Sc models often assume three facets (AEf) without orientation (FOn).

If the 31 different classes are considered, the best performance during the daytime for Q^* is from the FOo class (median RMSE of 43 W m^{-2}). There are two classes with an absolute median MBE of ≤3 W m^{-2} (L1, AE1). There are six models with both these characteristics (Fig. 3). For daytime Q_H, there are four classes with an MBE of <20 W m^{-2} (Vs, Lm, FO1, FOi). There is only one model with all of these (viz., Vs, Lm, FOi). The best overall performance for daytime Q_H, based on median RMSE, has a value of 69 W m^{-2} (FO1), but there are seven other classes within 4 W m^{-2} of this (Vs, Vi, Ls, Lm, R1, AE1, Sr) and three additional classes within 7 W m^{-2} (ANm, FOo, FOi), thereby accounting for all seven major classes (Fig. 2). No models have all of these characteristics, while two have five of them but do not generally fall within the group of best-performing models.

At night, best performance for Q^* is associated with ANm, Lm, Ri (median RMSE 11–13 W m^{-2} and/or median MBE < |4| W m^{-2}) and for Q_H with Vi, Lm, and FOi (median MBE = −2, −1, and −6, median RMSE = 14, 27, and 27 W m^{-2}, respectively). The Sr and Sc models have a similarly good RMSE (17–18 W m^{-2}).

For daytime Q_E, best overall performance is from Ls, FO1, AE1, Vi, FOi, and R1 (median MBE < |10| W m^{-2}). For daytime ΔQ_S, models with median RMSE < 96 W m^{-2} are Vi, Ls, FO1, AE1, and Sr but based on the absolute MBE, the best-performing models are FOo (median MBE = ≤|10| W m^{-2}), and FO1, AE1, Sr, Vi, Lm, and Ls (median MBE < |30| W m^{-2}). At night, Lm, ANm, FOi, Sr, and Vi models perform well based on median MBE and RMSE (<|4| and/or <22 W m^{-2}).

6. Conclusions

Urban surface–atmospheric exchanges are modeled for a wide variety of applications. The large set of models, examined here, have a range of approaches, complexities, and parameter requirements. Through the first stage of the first international model comparison reported here, significant model developments have taken place and improvements in model performance have resulted.

Evaluation of 33 models, with Vancouver (VL92) data, shows that generally models have best overall capability to model Q^* and least capability to model Q_E (order Q^*, ΔQ_S, Q_H, and Q_E; Table 4). No model performs best or worst for all fluxes. In particular, it seems to be difficult to minimize both Q^* and Q_H errors. There is evidence that some classes of models perform better for individual fluxes but not overall. Typically, those that perform best during daytime do not perform best at night.

The daytime RMSE for Q_H is larger than for Q^* for all but four models. These four are characterized as having amongst the four largest Q^* RMSE values. For RMSE$_S$, there is the tendency for Q_H errors to be greater than for Q^*, although there are more cases where the errors are similar. The unsystematic errors are generally smaller than systematic errors, particularly for the most poorly performing models. For most models, Q_H has a positive MBE, which observational errors may contribute to.

Seven characteristics (relating to vegetation, Q_F, morphology, facets and orientations, reflection, albedo and emissivity, and ΔQ_S) are used to classify each model. Some of the greatest differences in model performance are found between classes of model that treat vegetation and reflections differently. Some of the smallest differences relate to approaches used to calculate the ΔQ_S followed by urban morphology. Not including vegetation, even at a site with limited vegetation, yields the poorest performance for all fluxes during the day (in terms of RMSE) and for Q_E at night. During the day, median RMSE for models that do not include Q_F is similar (or better) than for those that do. However, at night, median RMSE for models, which include Q_F shows better performance for Q^*, Q_H, and ΔQ_S. Models that account for urban morphology orientation, and also intersections, often have slightly better performance than schemes that do not (e.g., Q_H in the daytime). The addition of intersections, however, does not always improve performance appreciably and in some cases has a negative impact on model performance.

The results for reflection schemes vary between day and night and with statistical measure (RMSE or MBE). In general, using a bulk albedo–emissivity results in better performance for all fluxes during the day. Classifying based on method used to calculate ΔQ_S has the smallest difference in the median of the RMSE and MBE of all classes. The residual method performs better during the day for all fluxes, while at night, differences are less significant. Class combinations show no models display all characteristics associated with strongest performance, although two display a large proportion of these. In general, the simpler models perform as well as the more complex models based on all statistical measures.

These results are based on a short time series for one urban location. In phase 2, the same models will be evaluated using a second dataset (Grimmond et al. 2009b). These results raise a number of questions that will be considered, with different flux partitioning, a wider range of conditions, and a longer time series. Of particular interest is whether the same models and classes perform well; whether the relative ability to model the individual fluxes remain the same; and whether it is possible for any class of models to minimize errors in both Q^* and Q_H.

Acknowledgments. Funded (Grimmond) by the Met Office (P001550), European Union Framework 7 (7 FP7-ENV-2007-1): MEGAPOLI (212520), BRIDGE (211345). This is a contribution to COST728 (Enhancing mesoscale meteorological modeling capabilities for air pollution and dispersion applications). We thank all involved in the original dataset collection (especially James Voogt and Tim Oke) and their funding agencies. Funding from BSIK-COM29 (Steeneveld), CATER 2006-2202 (Baik), and other agencies that support all the time contributed by participating groups is acknowledged. Color versions of the figures can be obtained by e-mailing corresponding author Sue Grimmond.

REFERENCES

Best, M. J., 1998: Representing urban areas in numerical weather prediction models. *Proc. Second Urban Environment Symp.,* Albuquerque, NM, Amer. Meteor. Soc., paper 10A.2.

——, 2005: Representing urban areas within operational numerical weather prediction models. *Bound.-Layer Meteor.,* **114,** 91–109.

——, C. S. B. Grimmond, and M. G. Villani, 2006: Evaluation of the urban tile in MOSES using surface energy balance observations. *Bound.-Layer Meteor.,* **118,** 503–525.

Chen, F., and J. Dudhia, 2001: Coupling an 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.

——, H. Kusaka, M. Tewari, J. Bao, and H. Hirakuchi, 2004: Utilizing the coupled WRF/LSM/Urban modeling system with detailed urban classification to simulate the urban heat island phenomena over the Greater Houston area. *Proc. Fifth Conference on Urban Environment,* Vancouver, BC, Canada, Amer. Meteor. Soc., paper 9.11.

Ching, J., and Coauthors, 2009: National Urban Database and Access Portal Tool. *Bull. Amer. Meteor. Soc.,* **90,** 1157–1168.

Clarke, J. A., 1985: *Energy Simulation in Building Design*. Adam Hilger, 362 pp.

Dandou, A., M. Tombrou, E. Akylas, N. Soulakellis, and E. Bossioli, 2005: Development and evaluation of an urban parameterization scheme in the Penn State/NCAR Mesoscale model (MM5). *J. Geophys. Res.,* **110,** D10102, doi:10.1029/2004JD005192.

Dupont, S., and P. G. Mestayer, 2006: Parameterization of the urban energy budget with the Submesoscale Soil Model. *J. Appl. Meteor. Climatol.,* **45,** 1744–1765.

——, ——, E. Guilloteau, E. Berthier, and H. Andrieu, 2006: Parameterization of the urban water budget with the Submesoscale Soil Model. *J. Appl. Meteor. Climatol.,* **45,** 624–648.

Essery, R. L. H., and D. B. Clark, 2003: Developments in the MOSES 2 land-surface model for PILPS 2e. *Global Planet. Change,* **38,** 161–164.

——, M. J. Best, R. A. Betts, P. M. Cox, and C. M. Taylor, 2003: Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *J. Hydrometeor.,* **4,** 530–543.

Feddema, J. J., K. W. Oleson, G. B. Bonan, L. O. Mearns, L. E. Buja, G. A. Meehl, and W. M. Washington, 2005: The importance of land-cover change in simulating future climates. *Science,* **310,** 1674–1678.

Fortuniak, K., B. Offerle, and C. S. B. Grimmond, 2004: Slab surface energy balance scheme and its application to parameterisation of the energy fluxes on urban areas. *NATO ASI,* Kiev, Ukraine. [Available online at www.met.rdg.ac.uk/urb_met/NATO_ASI/talks.html.]

——, ——, and ——, 2005: Application of a slab surface energy balance model to determine surface parameters for urban areas, Lund eRep. *Phys. Geogr.,* **5,** 90–91.

Grimmond, C. S. B., 1992: The suburban energy balance: methodological considerations and results for a mid-latitude west coast city under winter and spring conditions. *Int. J. Climatol.,* **12,** 481–497.

——, 2006: Progress in measuring and observing the urban atmosphere. *Theor. Appl. Climatol.,* **84,** 3–22.

——, and T. R. Oke, 1986: Urban water balance II: Results from a suburb of Vancouver, B.C. *Water Resour. Res.,* **22,** 1404–1412.

——, and ——, 1991: An evaporation–interception model for urban areas. *Water Resour. Res.,* **27,** 1739–1755.

——, and ——, 1999a: Heat storage in urban areas: Observations and evaluation of a simple model. *J. Appl. Meteor.,* **38,** 922–940.

——, and ——, 1999b: Aerodynamic properties of urban areas derived from analysis of surface form. *J. Appl. Meteor.,* **38,** 1262–1292.

——, and ——, 1999c: Rates of evaporation in urban areas. Impacts of urban growth on surface and ground waters. International Association of Hydrological Sciences Publication 259, 235–243.

——, and ——, 2002: Turbulent heat fluxes in urban areas: Observations and local-scale urban meteorological parameterization scheme (LUMPS). *J. Appl. Meteor.,* **41,** 792–810.

——, ——, and D. G. Steyn, 1986: Urban water balance I: A model for daily totals. *Water Resour. Res.,* **22,** 1404–1412.

——, H. A. Cleugh, and T. R. Oke, 1991: An objective urban heat storage model and its comparison with other schemes. *Atmos. Environ.,* **25B,** 311–326.

——, C. Souch, and M. Hubble, 1996: The influence of tree cover on summertime energy balance fluxes, San Gabriel Valley, Los Angeles. *Climate Res.,* **6,** 45–57.

——, J. A. Salmond, T. R. Oke, B. Offerle, and A. Lemonsu, 2004: Flux and turbulence measurements at a dense urban site in Marseille: Heat, mass (water, carbon dioxide) and momentum. *J. Geophys. Res.,* **109,** D24101, doi:10.1029/2004JD004936.

——, and Coauthors, 2009a: Urban Surface Energy Balance Models: Model characteristics and methodology for a comparison study. *Meteorological and Air Quality Models for Urban Areas,* A. Baklanov et al., Eds., Springer-Verlag, 97–124.

——, and Coauthors, 2009b: The International Urban Energy Balance Comparison Project: Initial results from phase 2. *Proc. Seventh Int. Conf. on Urban Climate,* Yokohama, Japan, International Association for Urban Climate paper A11-6, 4 pp. [Available online at http://www.ide.titech.ac.jp/~icuc7/extended_abstracts/pdf/372811-1-090524190750-003.pdf.]

Guilloteau, E., 1998: Optimized computation of transfer coefficients in surface layer with different momentum and heat roughness length. *Bound.-Layer Meteor.,* **87,** 147–160.

Hamdi, R., and G. Schayes, 2007: Validation of Martilli's urban boundary layer scheme with measurements from two mid-latitude European cities. *Atmos. Chem. Phys.,* **7,** 4513–4526.

——, and V. Masson, 2008: Inclusion of a drag approach in the Town Energy Balance (TEB) scheme: Offline 1D evaluation in a street canyon. *J. Appl. Meteor. Climatol.,* **47,** 2627–2644.

Hanna, S. R., and J. C. Chang, 1992: Boundary layer parameterizations for applied dispersion modeling over urban areas. *Bound.-Layer Meteor.,* **58,** 229–259.

——, and ——, 1993: Hybrid Plume Dispersion Model (HPDM) improvements and testing at three field sites. *Atmos. Environ.,* **27A,** 1491–1508.

Harman, I. N., and S. E. Belcher, 2006: The surface energy balance and boundary layer over urban street canyons. *Quart. J. Roy. Meteor. Soc.,* **132,** 2749–2768.

——, M. J. Best, and S. E. Belcher, 2004a: Radiative exchange in an urban street canyon. *Bound.-Layer Meteor.,* **110,** 301–316.

——, J. F. Barlow, and S. E. Belcher, 2004b: Scalar fluxes from urban street canyons. Part II: Model. *Bound.-Layer Meteor.,* **113,** 387–410.

Heiple, S., and D. J. Sailor, 2008: Using building energy simulation and geospatial modelling techniques to determine high resolution building sector energy consumption profiles. *Energy Build.,* **40,** 1426–1436.

Henderson-Sellers, A., and R. E. Dickinson, 1992: Intercomparison of land-surface parameterizations launched. *Eos, Trans. Amer. Geophys. Union,* **73,** 195–196.

——, Z. L. Yang, and R. E. Dickinson, 1993: The Project for Intercomparison of Land-surface Parameterization Schemes. *Bull. Amer. Meteor. Soc.,* **74,** 1335–1349.

——, P. Irannejad, K. McGuffie, and A. Pitman, 2003: Predicting land-surface climates—better skill or moving targets? *Geophys. Res. Lett.,* **30,** 1777, doi:10.1029/2003GL017387.

Ichinose, T., K. Shimodozono, and K. Hanaki, 1999: Impact of anthropogenic heat on urban climate in Tokyo. *Atmos. Environ.,* **33,** 3897–3909.

Irranejad, P., A. Henderson-Sellers, and S. Sharmeen, 2003: Importance of land-surface parameterization for latent heat simulation in global atmospheric models. *Geophys. Res. Lett.,* **30,** 1904, doi:10.1029/2003GL018044.

Jacobson, M. Z., 1999: *Fundamentals of Atmospheric Modeling.* Cambridge University Press, 656 pp.

Kanda, M., T. Kawai, M. Kanega, R. Moriwaki, K. Narita, and A. Hagishima, 2005a: A simple energy balance model for regular building arrays. *Bound.-Layer Meteor.,* **116,** 423–443.

——, ——, and K. Nakagawa, 2005b: A simple theoretical radiation scheme for regular building arrays. *Bound.-Layer Meteor.,* **114,** 71–90.

——, M. Kanega, T. Kawai, H. Sugawara, and R. Moriwaki, 2007: Roughness lengths for momentum and heat derived from outdoor urban scale models. *J. Appl. Meteor. Climatol.,* **46,** 1067–1079.

Kawai, T., M. Kanda, K. Narita, and A. Hagishima, 2007: Validation of a numerical model for urban energy-exchange using outdoor scale-model measurements. *Int. J. Climatol.,* **27,** 1931–1942.

——, M. K. Ridwan, and M. Kanda, 2009: Evaluation of the Simple Urban Energy Balance Model using selected data from 1-yr flux observations at two cities. *J. Appl. Meteor. Climatol.,* **48,** 693–715.

Kawamoto, Y., and R. Ooka, 2006: Analysis of the radiation field at pedestrian level using a meso-scale meteorological model incorporating the urban canopy model. Preprints, *Sixth Int. Conf. on Urban Climate,* Göteburg, Sweden, WMO, 446–449.

——, and ——, 2009a: Accuracy validation of urban climate analysis model using MM5 incorporating a multi-layer urban canopy model. *Extended Abstracts, Seventh Int. Conf. on Urban Climate,* Yokohama, Japan, paper A10-1.

——, and ——, 2009b: Development of urban climate analysis model using MM5 Part 2 - Incorporating an urban canopy model to represent the effect of buildings (in Japanese). *J. Environ. Eng.,* **74,** 1009–1018.

Kikegawa, Y., Y. Genchi, H. Yoshikado, and H. Kondo, 2003: Development of a numerical simulation system toward comprehensive assessments of urban warming countermeasures including their impacts upon the urban buildings' energy-demands. *Appl. Energy,* **76,** 449–466.

Kondo, H., and F. H. Liu, 1998: A study on the urban thermal environment obtained through a one-dimensional urban canopy model (in Japanese). *J. Japan Soc. Atmos. Environ.,* **33,** 179–192.

——, Y. Genchi, Y. Kikegawa, Y. Ohashi, H. Yoshikado, and H. Komiyama, 2005: Development of a multi-layer urban canopy model for the analysis of energy consumption in a big city: Structure of the urban canopy model and its basic performance. *Bound.-Layer Meteor.,* **116,** 395–421.

Krayenhoff, E. S., and J. A. Voogt, 2007: A microscale three-dimensional urban energy balance model for studying surface temperatures. *Bound.-Layer Meteor.,* **123,** 433–461.

Kusaka, H., H. Kondo, Y. Kikegawa, and F. Kimura, 2001: A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models. *Bound.-Layer Meteor.,* **101,** 329–358.

Lee, S.-H., and S.-U. Park, 2008: A vegetated urban canopy model for meteorological and environmental modelling. *Bound.-Layer Meteor.,* **126,** 73–102.

——, C.-K. Song, J.-J. Baik, and S.-U. Park, 2009: Estimation of anthropogenic heat emission in the Gyeong-In region of Korea. *Theor. Appl. Climatol.,* **96,** 291–303.

Lemonsu, A., C. S. B. Grimmond, and V. Masson, 2004: Modelling the surface energy balance of an old Mediterranean city core. *J. Appl. Meteor.,* **43,** 312–327.

——, and Coauthors, 2008: Overview and first results of the Montreal Urban Snow Experiment 2005. *J. Appl. Meteor. Climatol.,* **47,** 59–75.

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.

Loridan, T., and Coauthors, 2010: Trade-offs and responsiveness of the single-layer urban canopy parametrization in WRF: An offline evaluation using the MOSCEM optimization algorithm and field observations. *Quart. J. Roy. Meteor. Soc.,* **136,** 997–1019.

Martilli, A., A. Clappier, and M. W. Rotach, 2002: An urban surface exchange parameterisation for mesoscale models. *Bound.-Layer Meteor.,* **104,** 261–304.

——, Y. A. Roulet, M. Junier, F. Kirchner, M. W. Rotach, and A. Clappier, 2003: On the impact of urban surface exchange parameterisations on air quality simulations: The Athens case. *Atmos. Environ.,* **37,** 4217–4231.

Masson, V., 2000: A physically-based scheme for the urban energy budget in atmospheric models. *Bound.-Layer Meteor.,* **41,** 1011–1026.

——, C. S. B. Grimmond, and T. R. Oke, 2002: Evaluation of the Town Energy Balance (TEB) scheme with direct measurements from dry districts in two cities. *J. Appl. Meteor.,* **41,** 1011–1026.

Mitchell, V. G., R. G. Mein, and T. A. McMahon, 2001: Modelling the urban water cycle. *Environ. Model. Software,* **16,** 615–629.

Moriwaki, R., M. Kanda, H. Senoo, A. Hagishima, and T. Kinouchi, 2008: Anthropogenic water vapor emissions in Tokyo. *Water Resour. Res.,* **44,** W11424, doi:10.1029/2007WR006624.

Noilhan, J., and J. F. Mahfouf, 1996: The ISBA land surface parameterisation scheme. *Global Planet. Change,* **13,** 145–159.

Offerle, B., C. S. B. Grimmond, and T. R. Oke, 2003: Parameterization of net all-wave radiation for urban areas. *J. Appl. Meteor.,* **42,** 1157–1173.

——, ——, and K. Fortuniak, 2005: Heat storage and anthropogenic heat flux in relation to the energy balance of a central European city center. *Int. J. Climatol.,* **25,** 1405–1419.

Oke, T. R., 1988: The urban energy balance. *Prog. Phys. Geogr.,* **12,** 471–508.

——, R. Spronken-Smith, E. Jauregui, and C. S. B. Grimmond, 1999: The energy balance of central Mexico City during the dry season. *Atmos. Environ.,* **33,** 3919–3930.

Oleson, K. W., G. B. Bonan, J. Feddema, M. Vertenstein, and C. S. B. Grimmond, 2008a: An urban parameterization for a global climate model. Part I: Formulation and evaluation for two cities. *J. Appl. Meteor. Climatol.,* **47,** 1038–1060.

——, ——, ——, and ——, 2008b: An urban parameterization for a global climate model. Part II: Sensitivity to input parameters and the simulated heat island in offline simulations. *J. Appl. Meteor. Climatol.,* **47,** 1061–1076.

Pigeon, G., D. Legain, P. Durand, and V. Masson, 2007: Anthropogenic heat release in an old European agglomeration (Toulouse, France). *Int. J. Climatol.,* **27,** 1969–1981.

Pitman, A. J., A. Henderson-Sellers, and Z.-L. Yang, 1990: Sensitivity of regional climates to localized precipitation in global models. *Nature,* **346,** 734–737.

Porson, A., I. N. Harman, S. I. Bohnenstengel, and S. E. Belcher, 2009: How many facets are needed to represent the surface energy balance of an urban area? *Bound.-Layer Meteor.,* **132,** 107–128.

Roth, M., 2000: Review of atmospheric turbulence over cities. *Quart. J. Roy. Meteor. Soc.,* **126,** 941–990.

Rowley, F. B., A. B. Algren, and J. L. Blackshaw, 1930: Surface conductances as affected by air velocity, temperature and character of surface. *ASHRAE Trans.,* **36,** 429–446.

Ryu, Y.-H., J.-J. Baik, and S.-H. Lee, 2009: A new single-layer urban canopy model for use in mesoscale atmospheric models. *Proc. Seventh Int. Conf. on Urban Climate,* Yokohama, Japan, International Association for Urban Climate, paper A16-1. [Available online at http://www.ide.titech.ac.jp/~icuc7/extended_abstracts/pdf/374405-1-081210150038-004%28rev%29.pdf.]

Sailor, D. J., and L. Lu, 2004: A top-down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas. *Atmos. Environ.,* **38,** 2737–2748.

Salamanca, F., and A. Martilli, 2009: A new Building Energy Model coupled with an Urban Canopy Parameterization for urban climate simulations—part II. Validation with one dimension off-line simulations. *Theor. Appl. Climatol.,* doi:10.1007/s00704-009-0143-8.

——, A. Krpo, A. Martilli, and A. Clappier, 2009: A new building energy model coupled with an urban canopy parameterization for urban climate simulations—part I. formulation, verification, and sensitivity analysis of the model. *Theor. Appl. Climatol.,* doi:10.1007/s00704-009-0142-9.

Shashua-Bar, L., and M. E. Hoffman, 2002: The green CTTC model for predicting the air temperature in small urban wooded sites. *Build. Environ.,* **37,** 1279–1288.

——, and ——, 2004: Quantitative evaluation of passive cooling of the UCL microclimate in hot regions in summer. *Build. Environ.,* **39,** 1087–1099.

Taylor, K., 2001: Summarizing multiple aspects of model performance in single diagram. *J. Geophys. Res.,* **106** (D7), 7183–7192.

Tremback, C. J., and R. Kessler, 1985: A surface temperature and moisture parameterization for use in mesoscale numerical models. *Proc. Seventh Conf. on Numerical Weather Prediction,* Montreal, Quebec, Amer. Meteor. Soc., 433–434.

UN, cited 2009: World urbanization prospects, The 2007 Revision Population Database. [Available online at http://esa.un.org/unup/.]

Voogt, J. A., and C. S. B. Grimmond, 2000: Modeling surface sensible heat flux using surface radiative temperatures in a simple urban area. *J. Climate Appl. Meteor.,* **39,** 1679–1699.

Willmott, C. J., 1981: On the validation of models. *Phys. Geogr.,* **2,** 184–194.

Wilson, M. F., A. Henderson-Sellers, R. E. Dickinson, and P. J. Kennedy, 1987: Sensitivity of the Biosphere–Atmosphere Transfer Scheme (BATS) to the inclusion of variable soil characteristics. *J. Climate Appl. Meteor.,* **26,** 341–362.

Zilitinkevich, S. S., 1995: Nonlocal 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.