Multi-site evaluation of an urban land-surface model: intra-urban heterogeneity, seasonality and parameter complexity requirements

Article

Published Version

RMetS
Royal Meteorological Society

Multi-site evaluation of an urban land-surface model: intra-urban heterogeneity, seasonality and parameter complexity requirements

Thomas Loridan* and C.S.B. Grimmond

Environmental Monitoring and Modelling Group, Department of Geography, King's College London, UK

*Correspondence to: T. Loridan, King's College London, Environmental Monitoring and Modelling Group, Department of Geography, King's College London, London WC2R 2LS, UK.
E-mail: thomas.loridan@gmail.com, sue.grimmond@kcl.ac.uk

An extensive off-line evaluation of the Noah/Single Layer Urban Canopy Model (Noah/SLUCM) urban land-surface model is presented using data from 15 sites to assess (1) the ability of the scheme to reproduce the surface energy balance observed in a range of urban environments, including seasonal changes, and (2) the impact of increasing complexity of input parameter information. Model performance is found to be most dependent on representation of vegetated surface area cover; refinement of other parameter values leads to smaller improvements. Model biases in net all-wave radiation and trade-offs between turbulent heat fluxes are highlighted using an optimization algorithm. Here we use the Urban Zones to characterize Energy partitioning (UZE) as the basis to assign default SLUCM parameter values. A methodology (FRAISE) to assign sites (or areas) to one of these categories based on surface characteristics is evaluated. Using three urban sites from the Basel Urban Boundary Layer Experiment (BUBBLE) dataset, an independent evaluation of the model performance with the parameter values representative of each class is performed. The scheme copes well with both seasonal changes in the surface characteristics and intra-urban heterogeneities in energy flux partitioning, with RMSE performance comparable to similar state-of-the-art models for all fluxes, sites and seasons. The potential of the methodology for high-resolution atmospheric modelling application using the Weather Research and Forecasting (WRF) model is highlighted. This analysis supports the recommendations that (1) three classes are appropriate to characterize the urban environment, and (2) that the parameter values identified should be adopted as default values in WRF. Copyright © 2011 Royal Meteorological Society

Key Words: urban energy exchanges; land-surface scheme; urban canopy models; model evaluation; Noah/SLUCM; parameter optimization; urban zones; FRAISE; UZE; WRF

Received 31 March 2011; Revised 5 October 2011; Accepted 6 October 2011; Published online in Wiley Online Library 14 November 2011

Citation: Loridan T, Grimmond CSB. 2012. Multi-site evaluation of an urban land-surface model: intra-urban heterogeneity, seasonality and parameter complexity requirements. *Q. J. R. Meteorol. Soc.* **138**: 1094–1113. DOI:10.1002/qj.963

1. Introduction

In recent decades the ability of numerical models to simulate atmospheric processes has improved significantly, whether designed for climate, air quality or weather prediction. Advances have resulted from improvements in the understanding of the relevant processes, along with methods to simulate them, and increases in computer resources. Integral to this has been the representation of surface–atmosphere exchanges (Mahfouf *et al.*, 1987; Chen

and Avissar, 1994a, 1994b). Inputs of energy and moisture from the Earth's surface to the lowest atmospheric level are particularly important as they are responsible for changes in the atmospheric state variables. Consequently, land-surface models (LSMs) have been integrated in atmospheric modelling systems to simulate ground-level sources and sinks controlling the lower boundary conditions for each model grid box. With increasing grid resolution, a high level of horizontal heterogeneity is inherent in the representation of the corresponding fluxes; grid cells characterizing forest, grass fields, water bodies or urban environments might all coexist in a typical model domain leading to very distinct partitioning of the outgoing energy between radiative, turbulent (sensible and latent) or storage heat fluxes.

Initially, urban areas were ignored and natural surfaces were of primary interest (e.g. Deardorff, 1978; Sellers *et al.*, 1986; Noilhan and Planton, 1989; Chen *et al.*, 1996). However, the finer grid resolutions have meant that urban areas may extend over multiple grid cells, and concomitantly the increased interest in cities has resulted in urban surface exchanges gaining greater attention. The increased roughness from buildings impacts atmospheric flow and enhances turbulent exchange, altering the sources/sinks of heat, moisture and momentum compared to natural surfaces. Other important attributes of surface exchanges from cities include: radiation trapping in street canyons; large heat storage in the urban fabric (providing night-time energy release); limited evaporative fluxes from a reduced vegetation coverage (increased Bowen ratio); and anthropogenic heat release. Parametrization schemes designed to model the energy balance of such environments have flourished in the last decade (see references in Grimmond *et al.*, 2010).

With typical configurations currently used in numerical weather prediction (NWP), intra-urban heterogeneities are of importance, and contrasts in, for instance, typical street canyon morphology or building density from neighbouring grid cells need to be accounted for as they impact energy partitioning patterns (Loridan and Grimmond, 2011). Here we present an extensive evaluation of the Noah/Single Layer Urban Canopy Model (Noah/SLUCM: Chen and Dudhia, 2001; Kusaka *et al.*, 2001; Kusaka and Kimura, 2004) implemented in the v3.2 release of the community Weather Research and Forecasting (WRF) modelling system (Skamarock *et al.*, 2008).

In WRF, three default sets of urban input parameters are available for use with the National Land Cover Data (NLCD 1992) categories (Chen *et al.*, 2011). WRF users can assign urban cells to one of these three categories and the corresponding parameter values are used to run the SLUCM. Several other urban classifications have been developed for different purposes; examples include the Urban Climate Zones (UCZ: Oke, 2004) and Thermal Climate Zones (TCZ: Stewart and Oke, 2009) designed to characterize urban heat-island observations, or the urban terrain zones (Ellefsen, 1990) which characterize morphological differences of North American cities. Here we use two classifications designed specifically to study energy exchanges: (1) the urban categories from the Jackson *et al.* (2010) database (section 2.3), and (2) the Urban Zones to characterize Energy partitioning (UZE: Loridan and Grimmond, 2011; see section 2.5). The main aim of the study is to provide the SLUCM parameter values for the UZE categories and to evaluate the performance against a wide range of heat flux observations.

The study provides a detailed assessment of (1) the ability of the Noah/SLUCM to reproduce the surface energy balance observed in a range of urban environments, including seasonal changes, and (2) the impact of increasing effort or complexity to obtain input parameter data. It is conducted in five stages (Table I) using the methodology described in section 2. This involves observational data from 15 sites worldwide: North America (10, 7 in the USA), Europe (3), Australia (1) and Africa (1). For three cases, forcing and evaluation data are available for more than a year (Melbourne, one year; Helsinki and Łódź, two years). The locations selected ensure a range of latitudinal, climatological and meteorological conditions, building styles and population densities. An improved set of parameter values, for the UZE classes identified, is derived as part of the results section (section 3) and we recommend their use instead of the current default values in WRF. These new parameters and the objective method to classify urban sites into a UZE are independently evaluated using data from the Basel Urban Boundary Layer Experiment (BUBBLE: Christen and Vogt, 2004; Rotach *et al.*, 2004) and compared to current WRF settings.

2. Model evaluation: methodology

In this study (Table I) Noah/SLUCM is run off-line using observed forcing data (incoming short- and long-wave radiation ($K\downarrow$ and $L\downarrow$), air temperature, relative humidity, wind components, pressure and precipitation) at sites where observed fluxes are available for evaluation (Table II). As incoming long-wave radiation was not observed for some sites, this was estimated from cloud fraction (NCDC database, 2009) along with measurements of air temperature and relative humidity (Loridan *et al.*, 2011). Forcing data at hourly and 30-minute intervals (Ouagadougou, Helsinki and Melbourne) were linearly interpolated to the 10 minutes required for the Noah/SLUCM run time step. The modelled fluxes for the first 10 minutes of each hour are used for evaluation against the hourly observations. The observed fluxes of turbulent sensible heat (Q_H) and net all-wave radiation (Q^*) are adjusted, as described in Loridan and Grimmond (2011), to account for anthropogenic heat contribution (Q_F) using an estimate from the globally applicable Large-scale Urban Consumption of energY model (LUCY: Allen *et al.*, 2011) at each location. The size of the Q_F contribution for each site is reported in Loridan and Grimmond (2011; their Table II). The mean midday (± 3 h from solar noon) Q_F values range from 0 (Ou03) to $\sim$85 W m^{-2} (Vancouver sites) with potential error quantified as $<$20 W m^{-2}. No anthropogenic contribution is considered for the turbulent latent heat flux (Q_E). This adjustment to the observed Q^* and Q_H is used to eliminate the direct modelling of Q_F which is done as a specified profile with 25 parameters by the SLUCM (Miao *et al.*, 2009). Here the focus is on the ability of the scheme to simulate Q^*, Q_H and Q_E, as they are key to the correct estimation of lower boundary conditions in atmospheric models.

More observational data are available for short summertime periods (Table II). However, we know seasonality impacts energy partitioning (Loridan and Grimmond, 2011) and NWP is required for all seasons. Noah accounts for seasonal changes in vegetation (evolution of leaf area, albedo

Table I. Definition of stages used in the model evaluation. References and indication of the complexity involved in parameter estimation are given for each stage.

Stage		Input parameter value specification	SLUCM version	References	Level of complexity
0	a	Same for all sites: corresponding to the 'High Intensity Residential' class (urban class 2, default in WRF)	as originally implemented in WRF (v2.2), see Table II	Kusaka *et al.* (2001) Kusaka and Kimura (2004)	low
	b	Three classes: Each site is assigned values from the most appropriate class (based on f_{urb})		Tewari *et al.* (2006)	medium
	c	Site-specific f_{urb} is used instead of class value			medium
1	a	As for Stage 0a (HIR) but with updated values (Table I from Chen *et al.*, 2011)	after modification of input parameter table (WRF v3.1 and after)	Loridan *et al.* (2010) Chen *et al.* (2011, Table I)	low
	b	As for Stage 0b (three classes)			medium
	c	As for Stage 0c (site-specific f_{urb})			medium
2		From parameter database; values assigned based on: (1) geographical region and (2) level of urbanization (subclasses)	as for Stage 1	Jackson *et al.* (2010)	medium/high
3		Site-specific (where possible) (i.e. from observation)	as for Stage 1	See Table I from Loridan and Grimmond (2011)	high
4	i	Site-specific f_{urb}; other parameters optimized (MOSCEM)	as for Stage 1	MOSCEM: Vrugt *et al.* (2003)	high
		All parameters from model optimization (MOSCEM)	as for Stage 1		high
5	a	One class only: MD. Default values updated to account for Stage 1b, 2, 3 and 4 results	as for Stage 1	Current study + Loridan and Grimmond (2011)	low
	b	Three classes (UZE) : HD/MD/LD based on Loridan and Grimmond (2011)			medium
	c	Site-specific f_{urb} is used instead of class value			medium

and roughness) and the SLUCM deals with shading patterns in street canyons. Hence the model should be able to reproduce changes in observed flux partitioning through the year. The long-term datasets (Mb06, ŁÓ01/02, HE08/09) are split into three periods for each year (e.g. HE8J: 'January–April', HE8M: 'May–August', HE8S: 'September–December'). This results in a total of 27 evaluation periods.

All state variables are initialized as in Loridan *et al.* (2010). The soil is assumed to be 'clay-loam' (Chen and Dudhia, 2001, their Table II) with an initial soil moisture content corresponding to field capacity. Although this may be high, it is likely reasonable if irrigation is common, which is the case where many of these data were collected. The initial soil temperature profile (T_{soil}) is derived for different depths ($z = 0.1$, 0.3, 0.6 and 1.0 m) following Sellers (1972):

$$T_{soil}(z,t) = T_{mean} + \Delta T_0 exp\left(-z\sqrt{\frac{\omega_{day}}{2k_d}}\right) sin\left(wt - \sqrt{\frac{\omega_{day}}{2k_d}}z\right).$$

The runs are initialized at midnight ($t = 0$), k_d is the soil diffusivity, set to 0.002 cm^2 s^{-1} and ω_{day} is the angular frequency of oscillations (h^{-1}). Following Sellers (1972; p. 136) it is assumed that the typical daily temperature cycle penetrates to a depth of 0.2–0.8 m (i.e. first three layers); ΔT_0 is the amplitude of the temperature wave, which we estimate here using the forcing data at each site, taking half the amplitude of the air temperature over the first 48 h. We consider monthly mean air temperature to be characteristic of the variations at the lowest layer (1.0 m). To estimate T_{mean}, the monthly mean normal air temperatures from nearby sites are used (e.g. 'Clim81' from the NCDC database (2009), for all US sites). All facet temperatures (roof, wall and roads) are initialized to air temperature as available from the forcing data. Initially vegetation is set to the default class used in urban applications of WRF: 'cropland/grassland mosaic'. To provide initial 'spin-up', evaluation is conducted after the model has run for 48 h (starting at midnight). This methodology is designed to provide consistent initialization conditions across the sites database. For some of the campaigns, unusual conditions compared to normal (e.g. drought associated with irrigation bans in Vancouver) will not be accounted for in the evaluation and the model performance will be expected to suffer from it.

Table II. Category assigned to each site when considering the different classification methodology used in the study (NLCD 1992, 2001; Jackson *et al.*, 2010; UZE, Loridan and Grimmond, 2011).

Code	Site	Ref.	Period	f_{urb}	Z_R (m)	Substages 0b/1b classes (NLCD 1992)	NLCD 2001 classes (for comparison	Stage 2 classes (Jackson only)	Stage 4/5b classes (UZE: Loridan and Grimmond, 2011)
AR93**†	Arcadia, CA, USA	GO95	Jul–Aug 93	0.40	5.2	LIR	DLI	SW USA – LD	LD
AR94**†	Arcadia, CA, USA	GR96	Jul 94	0.43	5.2	LIR	DLI	SW USA – LD	LD
CH92**†	Chicago, IL, USA	GR94	Jul 92	0.55	6.7	LIR	DMI	NE USA – MD	MD
C95U**†	Chicago, IL, USA	KG97	Jun–Aug 95	0.61	5.9	LIR	DMI	NE USA – MD	MD
MA01	Marseille, France	GR04a	Jun–Jul 01	0.86	15.6	HIR	DHI	W Europe – MD	HD
ME93†	Mexico City, Mexico	OK99	Dec 93	0.98	18.4	C/I	DHI	Middle America – HD	HD
MI95	Miami, FL, USA	NE07	May–Jun 95	0.64	8.0	LIR	DHI	SE USA – MD	HD
OK03	Oklahoma City (Wood House), OK, USA	GR04b	Jun–Aug 03	0.22	4.5	LIR	DLI	SC USA – LD	LD
OU03†	Ouagadougou, Burkina Faso	OF05	Feb 03	0.9	3.0	HIR	DHI	W Africa – MD	HD
SG94†	San Gabriel, CA, USA	GR96	Jul 94	0.6	4.7	LIR	DMI	SW USA – LD	MD
VL92**	Vancouver, Canada	GO99	Aug 92	0.95	5.8	C/I	DHI	Canada – MD	HD
VS92**†	Vancouver, Canada	GO99	Jul/Sep 92	0.54	4.7	LIR	DMI	Canada – MD	HD
Mb06	Melbourne, Australia	CO07	Jan–Dec 06	0.63	6.4	LIR	DMI	Australia – MD	MD*
HE08/HE09	Helsinki, Finland	VE08	Jan 08–Dec 09	0.54	20.0	LIR	DMI	N Europe – MD	LD*
ŁÓ01/ŁÓ02	Łódź, Poland	OF06	Mar 01–Dec 02	0.7	10.6	LIR	DMI	E Europe – MD	HD*

Mean building height (Z_R) and plan-area fraction urban (f_{urb}). References for study details: GO95 (Grimmond and Oke, 1995), GR96 (Grimmond *et al.*, 1996), GR94 (Grimmond *et al.*, 1994), KG97 (King and Grimmond, 1997), GR04a (Grimmond *et al.*, 2004a), OK99 (Oke *et al.*, 1999), NE07 (Newton, 2007), GR04b (Grimmond *et al.*, 2004b), OF05 (Offerle *et al.*, 2005), GR96 (Grimmond *et al.*, 1996), GO99 (Grimmond and Oke, 1999), GR93 (Grimmond *et al.*, 1993), CO07 (Coutts *et al.*, 2007a, 2007b), VE08 (Vesala *et al.*, 2008), OF06 (Offerle *et al.*, 2006).
* Summer months are used to characterize the year-long campaigns at Stages 4/5. Note these are split into seasonal datasets for analysis.
** Measurement sites are in the same city but at different locations.
†Sites for which $L\downarrow$ needed to be modelled following Loridan *et al.* (2011).

2.1. *Stage 0: WRF v2.2 classes*

The Noah/SLUCM as initially implemented in WRF (v2.2: Tewari *et al.*, 2006) is run as a reference. It has three urban classes based on the US NLCD 1992 (Chen *et al.*, 2011): (1) low-intensity residential (LIR; where impervious surfaces account for 30–80% of total cover), (2) high-intensity residential (HIR; impervious surfaces 80–100% of the total cover), and (3) commercial/industrial (C/I; others not classified as HIR).

The 28 input parameters for the SLUCM (Table III) are the default values from WRF v2.2. These are directly linked to the NLCD 1992 definitions (see default class values of urban fraction: Table III). For turbulent exchanges the Jurges' relations are used instead of Monin–Obukhov formulations for wall and road surfaces (default setting in WRF), reducing the inputs required (see Table I, footnote *a*, from Chen *et al.* (2011), and Kusaka and Kimura (2004) for details on the parametrization).

Initially the HIR class is used for all sites, as this is the default urban class in WRF when no further information is provided by the user (Stage 0a, Table I). Each site is then assigned the most representative of the three classes (refined, Stage 0b) based on the observed vegetation fraction cover of the measurement footprint and NLCD 1992 definitions (Table II). Following NLCD 1992 definitions, 11 out of

Table III. Stage 0 input parameter list and corresponding default values, as implemented in WRF v2.2 (Tewari *et al.*, 2006), for 'low intensity residential' (LIR), 'high intensity residential' (HIR) and 'commercial/industrial' (C/I) NLCD 1992 classes.

Input parameters	NLCD 1992 land cover classes		
Definition	LIR	HIR	C/I
Urban fraction (−)	0.5	0.9	0.95
Roof height (m)	5	7.5	10
Roughness length above canyon for:			
momentum (m)	0.5	0.75	1.0
heat (m)	0.5	0.75	1.0
Zero plane displacement height (m)	1.0	1.5	2.0
Sky view factor (−)	0.62	0.56	0.48
Building coverage ratio (−)	0.5	0.5	0.5
Normalized road width (−)	0.5	0.5	0.5
Normalized building height (−)	0.3	0.4	0.5
Roughness length above roof for:			
momentum (m)	0.1	0.1	0.1
heat (m)	0.1	0.1	0.1
Drag coefficient by buildings (−)	0.1	0.1	0.1
Building volumetric parameter (m^{-1})	0.2	0.3	0.4
Albedo (−)			
roof	0.1	0.1	0.1
wall	0.1	0.1	0.1
road	0.1	0.1	0.1
Emissivity (−)			
roof	0.97	0.97	0.97
wall	0.97	0.97	0.97
road	0.97	0.97	0.97
Conductivity of materials (Cal cm^{-1} s^{-1} C^{-1})*			
roof	0.004	0.004	0.004
wall	0.004	0.004	0.004
road	0.004	0.004	0.004
Heat capacity of materials (Cal cm^{-3} C^{-1})*			
roof	0.5	0.5	0.5
wall	0.5	0.5	0.5
road	0.5	0.5	0.5
Total thickness of material layers (m)			
roof	0.2	0.2	0.2
wall	0.2	0.2	0.2
road	1.55	1.55	1.55

* Note non-SI units are given as they are in WRF v2.2; these differ from current WRF v3 ones (now in SI, Table IV).

the 15 sites used here are classified as LIR (Table II). The distinction between the other two categories (HIR and C/I) is rather subjective, and effectively it results in a two-class system, as default values are similar for these classes (Table III). For this study we assume the sites of VL92 and ME93 (i.e. the ones with the largest plan-area fraction built) belong to the C/I category, while MA01 and OU03 are from HIR. Although it is not used in this study because of a lack of consistency with the default class parameter values (Table III), the more recent NLCD 2001 provides more-objective categories which are all explicitly distinguished by impervious fraction cover (developed, low intensity (DLI): 20–49% of total cover; medium intensity (DMI): 50–79% of total cover; high intensity (DHI): 80–100% of total cover). Using these categories, three sites are classified as DLI, seven as DMI and five as DHI (Table II). It should be noted that for this sub-stage b, and similarly for Stages 1b, 4 and 5b, the plan-area urban fraction f_{urb} is the default for the class used (Table III) and not the site-specific f_{urb} value given in Table II. This has implications for the coupling of the SLUCM to Noah; in moving from Stage 0a to b the urban fraction is modified as well as the other urban parameters. Finally, a sub-stage c is performed where the default class value for f_{urb} is replaced by the site-specific estimate.

2.2. *Stage 1: WRF v3.0 classes*

In Stage 1 the modified SLUCM is used (Loridan *et al.*, 2010) with changes to the type and number of input parameters (Table IV). The reduction in parameters (cf. Table III) includes calculation of roughness parameters internally (Macdonald *et al.*, 1998; Kanda *et al.*, 2007; Loridan *et al.*,

2010, their Eqs 19 and 20). The default values (and units) of several parameters are also modified in WRF v3.0 (e.g. thermal conductivities, heat capacities, albedo and emissivity values: Chen *et al.*, 2011). Of the 30 urban parameters required by the SLUCM (Table III, Loridan *et al.*, 2010), 20 have their values modified in this evaluation depending on the stage considered. For all stages the empirical constant a_K, which relates the roughness length for heat to that of momentum (Kanda *et al.*, 2007) is fixed to 1.29; and the relative portion of material layers are kept at their default (but total material depth varies). The same sub-stages a (default), b (refined as most representative) and c (with site-specific f_{urb}) are considered as in Stage 0 (section 2.1). For sub-stages a and b the urban fraction is again the default for the class used (Table IV) and not the site-specific f_{urb} value given in Table II.

2.3. Stage 2: Jackson et al. (2010) parameter database

In Stage 2 the density/form of buildings and complexity of the site are accounted for in assigning parameter values. The underlying assumption is that cities from the same region share similar urban design and typical materials, which are determined by climatic conditions, available resources and cultural habits (Jackson *et al.*, 2010). The Jackson *et al.* (2010) database used in Stage 2 has four urban categories for each geographic region: Tall Building Districts (TBD) have buildings 10 storeys or more and a vegetation cover below 15%; High Density (HD) areas have buildings 3–10 storeys high and vegetation cover of 5 to 25%; Medium Density (MD) areas have row houses or apartment complexes 1–3 storeys tall with a vegetation cover of 20 to 60%; and Low Density (LD) areas have 1–2 storey buildings with a vegetation cover of 50 to 85%. For all 15 sites the geographical region and urban category (Table II) were used to extract parameter values from the Jackson *et al.* (2010) database. This database was designed for the urban scheme in the Community Land Model (CLM: Oleson *et al.*, 2008) and therefore provides most of the quantities required by the SLUCM (Table IV). As a consequence most values (including f_{urb}) vary for each site and the mean by category is reported in Table IV for comparison with other stages. The standard deviation of roof height (σ_Z) for each class, which cannot be directly extracted from the database, was estimated from the class description (Jackson *et al.*, 2010). Not unexpectedly, none of the evaluation sites belong to the TBD category as there is a relative scarcity of such sites and a difficulty in obtaining representative measurements in such an environment.

2.4. Stage 3: site-specific values

In Stage 3 site-specific (measured) values are assigned to parameters. Mean building geometry (Z_{R}, W_{road}, W_{roof}, Table IV) and surface fraction cover (f_{urb}) characteristic of the measurement footprints are well documented (Grimmond and Oke, 1999, their Table II; Grimmond and Oke, 2002, their Table III) but that is not the case for typical material properties. Model runs of Masson's (2000) Town Energy Balance (TEB) for some of the sites were used to derive the required inputs (e.g. MA01: Lemonsu *et al.*, 2004; ME93: Masson *et al.*, 2002; OK03 (wood house site): Anderson, 2009; ŁÓ01/ŁÓ02 and OU03: Offerle, 2003 and personal communication) and the international urban model energy balance intercomparison project documentation, hereafter termed 'PILPS-urban' (Grimmond *et al.*, 2010, 2011) were used for VL92 and Mb06. For the remaining sites (i.e. MI95, C95U, CH92, SG94, AR93/94 and HE08/09) the material property values from Stage 2 are used.

2.5. Stage 4: parameter optimization

In Stage 4 all input parameters are optimized using the Multi Objective Shuffled Complex Evolution Metropolis (MOSCEM: Vrugt *et al.*, 2003) algorithm. Before the site parameters are optimized, however, they are categorised using the Urban Zones to characterize Energy partitioning (UZE) method (Loridan and Grimmond, 2011). This is based on analysis of observed flux partitioning patterns and the fraction of the surface that is actively involved in energy exchanges (Loridan and Grimmond (2011) provide details; viz. their Eqs 19–24):

1. *Built index* (χ_{built}) is the ratio of the total built fabric (roof/road/walls) in direct solar radiation relative to the total three-dimensional (3-d) surface cover (i.e. including leaf/tree area). The value of χ_{built} varies with season and time of day.
2. *Vegetated index* (χ_{veg}) is the ratio of the vegetated surface (accounting for seasonal variation of the leaf area index calculated using Loridan *et al.* (2011), their Eq. 12) to the total 3-d surface cover.

The observed midday fluxes against surface indices were used to objectively identify that there are *three* classes (UZE) and that they have the following characteristics (Flux Ratio – Active Index Surface Exchange (FRAISE) scheme; see Loridan and Grimmond (2011) for full description) summarized by Figure 1:

1. High Density (HD) sites have $\chi_{\mathrm{built}} > 0.11$ *but* $\chi_{\mathrm{veg}} < 0.43$ and, on average, are expected to store a greater portion of the daytime incoming radiative energy ($Q{\downarrow} = K{\downarrow} + L{\downarrow}$, see Figure 1) than they dissipate via turbulent latent (Q_{E}) and sensible (Q_{H}) heat fluxes (see Fig. 11 in Loridan and Grimmond, 2011).
2. Medium Density (MD) sites have $\chi_{\mathrm{built}} > 0.11$ *and* $\chi_{\mathrm{veg}} > 0.43$; they dissipate more $Q{\downarrow}$ as Q_{H} than energy that is stored.
3. Low Density (LD) sites have $\chi_{\mathrm{urb}} < 0.11$; they are the least urbanized and dissipate a greater portion of $Q{\downarrow}$ as Q_{H} and Q_{E} than is stored.

To classify the sites with annual data, summer χ index values are used (i.e. at the time of maximum energy exposure). Note that a fourth UZE was identified in Loridan and Grimmond (2011), representing conditions where both indices are below their respective thresholds; only wintertime datasets belonged to this category and here we consider these cases covered by LD. When the sites were classified (Table II), consideration was given to conditions during the measurements (Fig. 11, Loridan and Grimmond, 2011). In two cases conditions of the observations warranted changes: firstly, observations at one site were conducted under extremely humid conditions with a lot of external water, leading to vegetation being more active (i.e. actual χ_{veg} higher than the one computed; C95U reclassified from

Table IV. Mean input parameter values assigned at Stages 0b–5b.

	Input parameters		1: Most urban category						2: Medium urban category						3: Low urban category					
		Definition	0b	1b	2**	3**	4	**5b**	0b	1b	2**	3**	4	**5b**	0b	1b	2**	3**	4	**5b**
1	f_{urb}	Urban fraction	0.95	0.95	0.72	0.80	0.72	**0.75**	0.9	0.9	0.66	0.6	0.55	**0.6**	0.5	0.5	0.5125	0.46	0.49	**0.5**
2	Z_R	Roof height (m)	10	10	19.3	9.5	19.8	**18**	7.5	7.5	14.4	5.93	19.641	**15**	5	5	9.75	9.1	3.1	**6**
3	W_{road}	Road width (m)	10	10	22.6	10.0	4.3	**8**	9.4	9.4	23.7	15.5	6.59	**10**	8.3	8.3	19.69	10.1	16.9	**15**
4	W_{roof}	Roof width (m)	10	10	28	12.6	23.8	**22**	9.4	9.4	22.4	10.9	22.66	**20**	8.3	8.3	10.6	8.7	5.00	**8**
5	σ_Z	Standard deviation of roof height (m)	–	4	3.75	*	2.5	**3**	–	3	2.7	*	1.467	**1.5**	–	1	1.875	*	1.23	**1.0**
		Albedo (−)																		
6	α_{roof}	roof	0.1	0.2	0.24	0.17	0.05	**0.1**	0.1	0.2	0.3	*	0.0579	**0.1**	0.1	0.2	0.2425	*	0.127	**0.15**
7	α_{wall}	wall	0.1	0.2	0.34	0.27	0.09	**0.1**	0.1	0.2	0.34	*	0.055	**0.1**	0.1	0.2	0.4625	*	0.092	**0.1**
8	α_{road}	road	0.1	0.2	0.13	0.12	0.25	**0.15**	0.1	0.2	0.13	*	0.219	**0.15**	0.1	0.2	0.13	*	0.05	**0.15**
		Emissivity (−)																		
9	ε_{roof}	roof	0.97	0.9	0.71	0.908	0.87	**0.85**	0.97	0.9	0.71	*	0.851	**0.85**	0.97	0.9	0.795	*	0.867	**0.85**
10	ε_{wall}	wall	0.97	0.9	0.89	0.89	0.96	**0.90**	0.97	0.9	0.89	*	0.968	**0.90**	0.97	0.9	0.8825	*	0.862	**0.90**
11	ε_{road}	road	0.97	0.95	0.91	0.9355	0.86	**0.95**	0.97	0.95	0.91	*	0.944	**0.95**	0.97	0.95	0.91	*	0.890	**0.95**
		Conductivity ($W\,m^{-1}\,K^{-1}$)																		
12	k_{roof}	roof	1.67472	0.67	7.305	1.669	0.49	**0.8**	1.67472	0.67	0.38	*	0.419	**0.4**	1.67472	0.67	0.4625	*	0.190	**0.4**
13	k_{wall}	wall	1.67472	0.67	1.028	0.7538	2.17	**1.0**	1.67472	0.67	1.032	*	1.979	**1.0**	1.67472	0.67	0.6275	*	1.94	**1.0**
14	k_{road}	road	1.67472	0.4004	1.115	0.9588	0.27	**0.8**	1.67472	0.4004	1.115	*	0.259	**0.8**	1.67472	0.4004	1.115	*	2.09	**0.8**
		Heat capacity ($J\,m^{-3}\,K^{-1}$)																		
15	C_{roof}	roof	2.09E6	1.0E6	1170704	1558000	2281730	**1.5E6**	2.0934E6	1.0E6	843485	*	2231110	**1.2E6**	2.0934E6	1.0E6	1072875	*	976634	**1.0E6**
16	C_{wall}	wall	2.09E6	1.0E6	953753	1518000	2200620	**1.4E6**	2.0934E6	1.0E6	826000	*	2277430	**1.2E6**	2.0934E6	1.0E6	659265	*	2131500	**1.2E6**
17	C_{road}	road	2.09E6	1.4E6	1886382	1586000	338202	**1.5E6**	2.0934E6	1.4E6	1886382	*	309253	**1.5E6**	2.0934E6	1.4E6	1886382	*	2290060	**1.5E6**
		Total thickness (m)																		
18	$d_{z,roof}$	roof	0.2	0.2	0.077	0.1518	0.44	**0.5**	0.2	0.2	0.114	*	0.458	**0.5**	0.2	0.2	0.0425	*	0.460	**0.5**
19	$d_{z,wall}$	wall	0.2	0.2	0.288	0.316	0.97	**0.3**	0.2	0.2	0.276	*	0.614	**0.3**	0.2	0.2	0.195	*	0.999	**0.3**
20	$d_{z,road}$	road	1.55	1.55	2.55	1.218	0.68	**1.0**	1.55	1.55	2.55	*	0.617	**1.0**	1.55	1.55	2.55	*	1.933	**1.0**
		Vegetation class	1	1	1	1	1	**1**	1	1	1	1	2	**2**	1	1	1	1	2	**2**

Parameter numbers are used in Figures 5(a), 6(a) and 7(a).
* Stage 2 values kept in Stage 3 when not enough information was available from literature.
** The classes used for the mean at Stages 2 and 3 are the UZEs. Vegetation classes are from Chen and Dudhia (2001): 1: cropland/grassland mosaic; 2: shrubland/grassland.

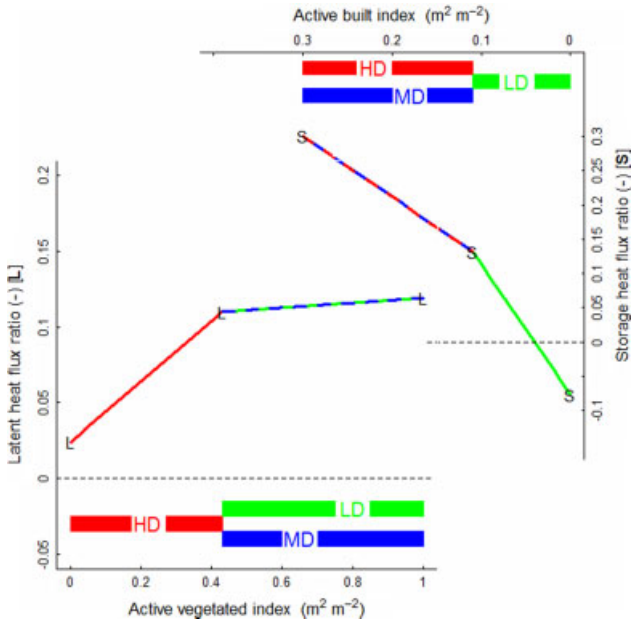

Figure 1. Mean midday (±3 hours around solar noon) summertime flux ratios (normalized by the total incoming short-wave and long-wave radiative energy: $Q\downarrow = K\downarrow + L\downarrow$) versus active surface indices (calculated with the FRAISE scheme (Loridan and Grimmond, 2011)). The range of active index values covered by the three UZE categories (LD/MD/HD) is shown. Note the active built index axis is reversed. See text for further details. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

category HD to MD); while, secondly, drought conditions resulted in a reduction of vegetation impact on energy exchange (i.e. χ_{veg} lower than computed; VS92 moved from category MD to HD).

To allow optimization of all input parameters for a given UZE using MOSCEM, a 'super dataset' of flux observations was created for each class. For each measurement campaign a week's data with minimal gaps is used, plus five additional days of forcing data for spin-up that are not used in the optimization. First, the observed urban fraction (f_{urb}) for each site (Table II) is used and all other parameters are optimized simultaneously (i.e. a multi-parameter optimization). SLUCM runs using these optimum parameter values and site-specific f_{urb} are referred to as Stage 4i (Table I) in the model performance analysis (section 3). The optimization windows (minimum and maximum thresholds) are the same as Loridan *et al.* (2010, their Table III) except for the building geometry parameters: $3\ \text{m} \leq Z_R \leq 20\ \text{m}$; $5\ \text{m} \leq W_{roof} \leq 25\ \text{m}$; $4\ \text{m} \leq W_{road} \leq 40\ \text{m}$. The initial, default values used are those assigned at Stage 1b (Table IV). As building height varies, the height of forcing level is adapted to maintain the same absolute distance between Z_R and the measurement height. As f_{urb} is fixed, the optimization will not impact simulation of evaporative fluxes so only two metrics ('objectives') are used here: the Root Mean Square Errors (RMSEs) for Q^* and Q_H. MOSCEM is set to optimize 1000 combinations of the parameters. As f_{urb} has a major impact on flux partitioning and can constrain model performance independent of other parameter values, the optimum f_{urb} value is determined with a second run with f_{urb} allowed to vary between 0 and 1 and all other parameters fixed to their optimum (as identified in the first run, Table IV). The three metrics for the second run are the RMSEs for Q^*, Q_H and Q_E, and 100 optimum f_{urb} values are used. The aim of this two-step procedure is to identify one parameter set per class that best represents (in terms of model performance) the sites within the class. The performance of the SLUCM using these is referred to as Stage 4.

2.6. *Stage 5: synthesis/recommendation*

Although the parameter estimates from the optimisation (Stage 4) should be the best, these do not necessarily contain physically reasonable combinations. MOSCEM will try to compensate for biases in the SLUCM by setting some parameters to their extreme values (Loridan *et al.*, 2010). Stage 5 is a synthesis/recommendation stage where results from Stages 1b, 2, 3 and 4 are revisited to derive a new set of parameter values for each UZE which (1) improves the performance of the scheme with regards to Stage 1 (current default in WRF), and (2) are physically realistic urban configurations. As for Stages 0 and 1, in sub-stage a all parameters are assigned the MD values. For sub-stage b, class values correspond to the appropriate UZE category of each site, whereas sub-stage c involves site-specific f_{urb}.

2.7. *Independent evaluation*

To independently assess the recommendations made here, data for three urban sites from BUBBLE (Christen and Vogt, 2004; Rotach *et al.*, 2004) are used. The three urban sites used are: BSPR: 'Basel-Sperrstrasse', dense urban; BSPA: 'Basel-Spalenring', dense urban; and ALLS: 'Allschwil-Ramelstrasse', suburban with data period analysed from the Intensive Observation Period (IOP, June–July 2002: Christen and Vogt, 2004). To objectively determine which UZE these sites should be assigned to, their active surface indices (χ_{built} and χ_{veg}, section 2.5) are calculated using the methods in Loridan and Grimmond (2011). The results from this assign the three sites into each of the three classes; this would match NLCD 2001 classes but not the NLCD 1992 ones (Table V). Off-line model runs are performed using the corresponding default parameter values from Stages 1b and 5b (Table IV) with similar procedures to those outlined in section 2. Forcing data available at 10 min intervals were used for the model runs (i.e. rather than interpolated from hourly or 30 min) and model outputs were averaged to hourly for evaluation.

2.8. *Evaluation statistics*

At each stage the evaluation is performed over 38 063 hours when observations are available. The model performance statistics used for each stage and dataset are the RMSEs. In addition, summary performance is provided by the Mean Bias Error (MBE) and Tukey's (1977) box plot or 'schematic plot' with the median, lower and upper quartiles (and the interquartile range, IQR) shown as well as whiskers for values within 1.5 IQR of the upper and lower quartiles (Wilks, 1995). The individual dots indicate 'outside' values (see Figure 3 for example).

To objectively characterize the overall performance of the scheme, at each stage the sum of RMSEs ($\text{RMSE}_{\sum}$) and MBEs ($\text{MBE}_{\sum} = \{\sum \text{MBE}^2\}^{0.5}$) over Q^*, Q_H and Q_E is used. The minimum $\text{RMSE}_{\sum}$ is considered indicative of the overall model performance as it combines the errors in energy release to the atmosphere, which is most important in atmospheric modelling applications. Depending on the

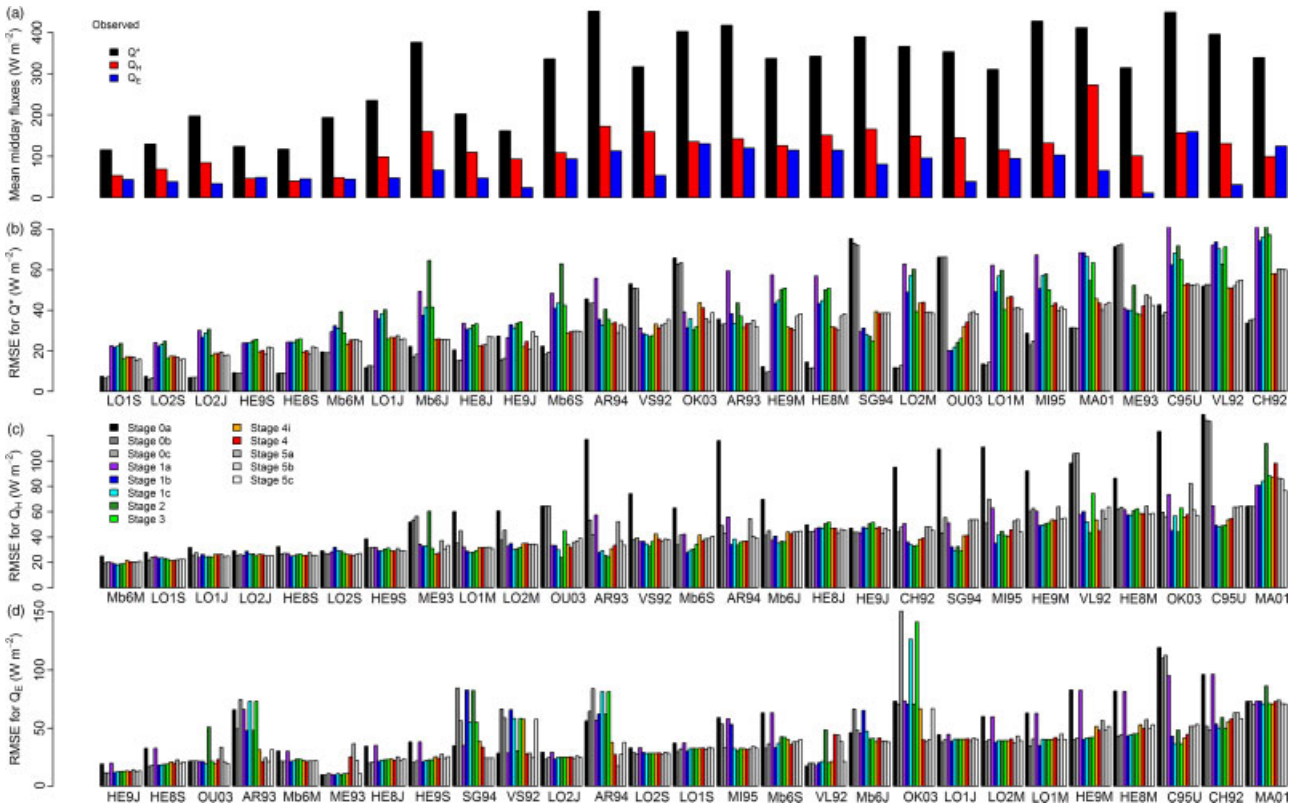

Figure 2. (a) Observed mean midday (±3 hours around solar noon) fluxes and RMSE model performance for (b) Q^*, (c) Q_H, and (d) Q_E for the 27 datasets (Table II) and different Stages (Table I). Note that for (a) the sites are ordered as in (b). This figure is available in colour online at wileyonlinelibrary.com/journal/qj

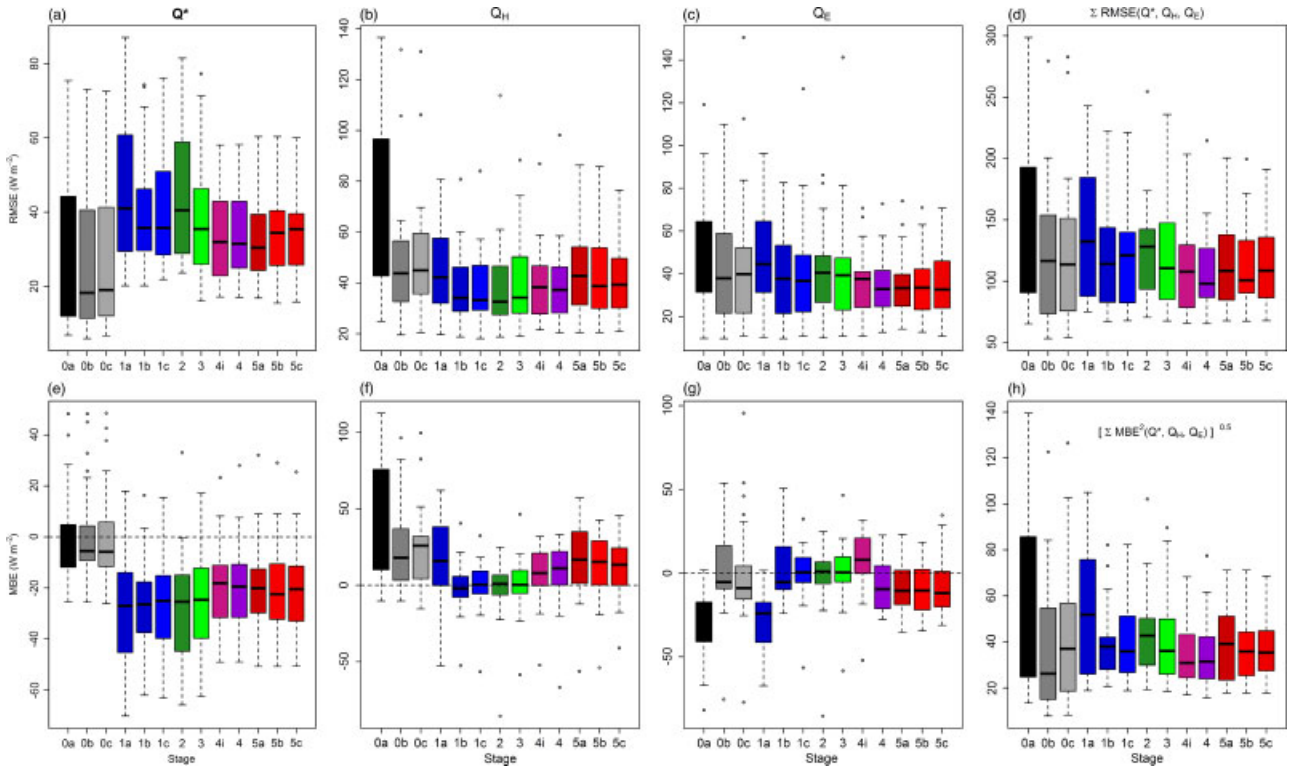

Figure 3. Tukey's (1977) schematic plots summarizing results from the 27 datasets shown in Figure 2: (a)–(c) RMSE for Q^*, Q_H and Q_E; (d) $\text{RMSE}_\Sigma = \Sigma\ \text{RMSE}(Q^*, Q_H, Q_E)$; (e)–(g) MBEs for Q^*, Q_H and Q_E; and (h) $\text{MBE}_\Sigma = \{\Sigma\ \text{MBE}^2(Q^*, Q_H, Q_E)\}^{0.5}$. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

application, though, a user might consider one of the fluxes to be more important (e.g. Q_H when considering boundary-layer processes); however, in the current study we do not focus on any specific application and consider each flux equally relevant.

3. Model evaluation: results

The model performance (section 2.8) for Q^*, Q_H and Q_E by stage and dataset, ordered by increasing RMSE at Stage 5b (Table I) for each flux for the 27 datasets, is presented in Figure 2. Summary box-plots are provided at each stage for RMSE (Figure 3(a)–(c)) plus overall performance ($RMSE_{\sum}$, Figure 3(d)) and similarly for MBE (Figure 3(e)–(h)).

3.1. Stages 0 and 1: WRF classes v2.2 and v3.0

The model changes introduced in Stage 1 (section 2.2) result in an obvious trade-off in performance relative to Stage 0. Simulation of Q_H greatly improves for almost all datasets (except MA01 and ŁÓ2S) but is generally associated with a decrease in the performance of Q^* (Figure 2(b) and (c), see for instance C95U or MI95). The MBE box-plots (Figure 3(e) and (f)) show an overestimation of Q_H and an underestimation of Q^* at Stage 0. At Stage 1 the bias in Q_H is reduced; however, this leads to an increase in the bias of Q^*. This systematic trade-off in the modelling of the two fluxes has already been noted for SLUCM (Loridan *et al.*, 2010) and more generally for the 33 models involved in Phase 2 of the PILPS-urban (Grimmond *et al.*, 2011). The reduction in sites with very large RMSEs for Q_H from Stage 0 to Stage 1 (Figure 2(c)) suggests more universal better performance. This is clear from the box-plots of Figure 3(b) and (f).

Given the Noah/SLUCM tile approach, where there is no interaction until passed to another grid level, the ability to simulate Q_E in Stage 0a (0b) is minimally affected when switching to 1a (1b), as Stage 1 modifications mainly relate to SLUCM. The small differences noticeable from Figure 3(c) and (g) (e.g. reduction of the IQR from 0b to 1b) are attributable to a change in the monthly varying roughness-length coefficients from Noah between WRF v2.2 and v3.0 releases. On the other hand, the reduction in RMSE from refining the site classification (a to b stage) is obvious (Figure 3(c)), with a decrease in MBE (Figure 3(g)). This highlights the critical role played by the urban fraction (f_{urb}, Tables III and IV) in the simulation of Q_E. This improvement in the partitioning of the turbulent fluxes is directly reflected in Q_H which is consistently improved from a to b with a median MBE value <1 W m^{-2} by Stage 1b (Figure 3(f)). The IQR and spread of results (length of the whiskers) are also both reduced considerably for the RMSE and MBE in sub-stage b (Figure 3(b) and (f)). For Q^* the performance is less influenced at sub-stage b (Figure 3(e)) but there is some improvement in RMSE, potentially from the refined characterization of canyon geometry (Tables III and IV). When further information is provided about the plan-area fraction vegetated (i.e. site-specific, sub-stage c), some improvement in the modelling of Q_E can be noticed at Stage 1 (reduced bias, Figure 3(g)) but the overall performance change is minimal (e.g. Figure 3(d)). This would suggest that although the benefit from a refinement of classes is immediate (i.e. sub-stage b) the Noah/SLUCM is not able to take full advantage of the most accurate estimation of the plan-area fraction vegetated. Stage 1b appears as the best option when the net $RMSE_{\sum}$ (Figure 3(d)) is considered, with similar median but lower IQR and smaller outliers than Stage 0b.

3.2. Stages 2 and 3: Jackson et al. (2010) database and site-specific values

In Stages 2 and 3 the complexity (and effort required) in parameter estimation is considerably increased (Table I) but does not result in significant improvement in the model performance: RMSEs and MBEs for Stages 2 and 3 remain similar to Stage 1b/1c, with a systematic underestimation of Q^* (Figure 3(e)) and a near-zero median MBE for Q_H (Figure 3(f)) and Q_E (Figure 3(g)). This suggests that even with site-specific values the scheme fails to simulate the magnitude of the three fluxes at the same time and highlights the trade-off issues already mentioned in section 3.1. In line with previous comments about sub-stages c (section 3.1), this also suggests that the scheme is not able to profit from the best available characterization of the surface, given model biases.

Considering the individual datasets (Figure 2), several anomalies seem to appear: Q^* in Stage 2 for Melbourne (Mb6J, Mb6M and Mb6S) suggests that the Jackson *et al.* (2010) database is inappropriate for the site. The primary criteria to distinguish between a Low or Medium Density (LD or MD) site (Table II) is the urban fraction, so Mb06 is categorized as MD. However, the actual geometry of the site ($Z_R = 6.4$ m; $W_{road} = 38.5$ m; $W_{roof} = 15.2$ m) is not typical for MD sites in Australia ($Z_R = 15.0$ m; $W_{road} = 25$ m; $W_{roof} = 30.6$ m: Jackson *et al.*, 2010) and such differences have a strong impact on the radiation budget through trapping of incoming energy. Mexico (ME93) Stage 2 performance is considerably poorer than in 1b or 3 for both Q^* and Q_H because of the extremely high values assigned to thermal conductivities of roof layers for sites in this region of the world in the Jackson *et al.* (2010) database. This anomaly can also be seen in the averaged k_{roof} value provided for category 1 in Table IV. In OK03, when observed f_{urb} values are used in Stage 3 there is a deterioration in performance in Q_E (and Q_H). This is the case with the smallest f_{urb} (0.22) so SLUCM plays a much smaller role than Noah. The default vegetation class used by WRF is selected to enhance evaporation and compensate for the lack of Q_E in the SLUCM (i.e. 'cropland/grassland mosaic', see Loridan *et al.* (2010) for further discussion on this). However, when the vegetation fraction is large this leads to overestimation of Q_E (and consequently underestimation of Q_H), so poor RMSEs (Figure 2, similarly AR93/94). This suggests that a vegetation class with less evaporation (e.g. 'mixed shrubland-grassland') should be used in low f_{urb} sites. The impact on the mean diurnal Q_E and Q_H when the urban parameters are the Stage 3 values and two different vegetation classes are used can be seen in Figure 4. The vegetation class with less evaporation shows a clear improvement for both fluxes. Note the impact on Q^* is minimal (not shown).

3.3. Stage 4: parameter optimization

The positive bias in Q_E, as a result of the choice of vegetation class discussed in section 3.2, was further investigated using a sensitivity test during the MOSCEM optimization (i.e. the two optimization runs described in section 2.5 were

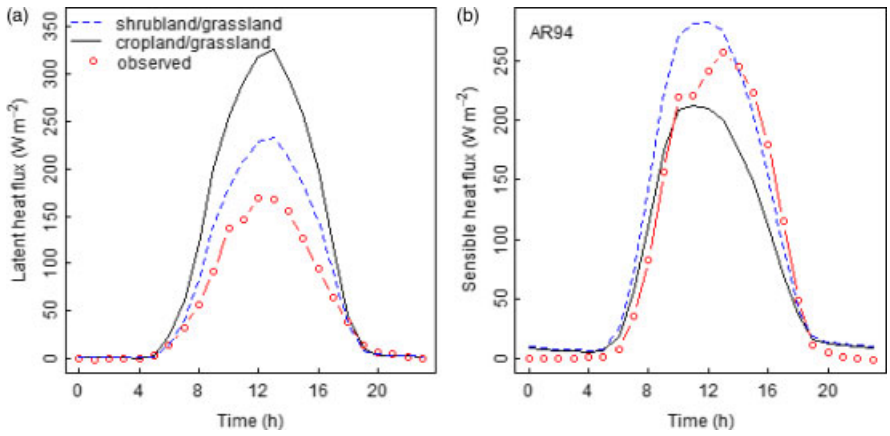

Figure 4. Observed (dots) and Noah/SLUCM simulated (Stage 3) mean diurnal (a) Q_E and (b) Q_H for the site of AR94 (Table II) when using the 'cropland/grassland' (solid line) and the 'shrubland/grassland' (dashed line) vegetation classes in Noah. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

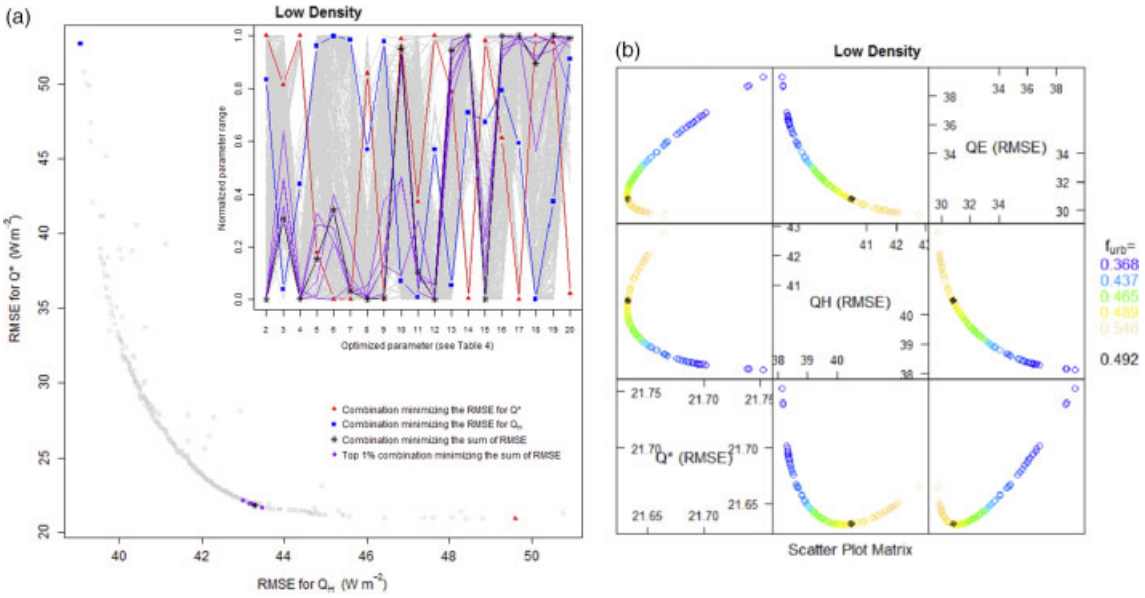

Figure 5. (a) RMSE performance for the 1000 optimum combinations identified by MOSCEM in the first step of the optimization procedure for LD sites. The normalized parameter values (parameters 2–20, Table IV) are shown (inset). The combinations minimizing the Q_H (blue square) and Q^* (red triangle) RMSEs are highlighted. The ten combinations leading to minimum $RMSE_\Sigma$ sum (purple dots) represent the best compromise (top 1%) among the 1000 solutions; the best option (black asterisk) is also shown. (b) RMSEs resulting from the triple-objective f_{urb} optimization ($0 < f_{urb} < 1$) with all other parameters set to their Stage 4 optimum (Table IV). The f_{urb} value leading to minimum $RMSE_\Sigma$ is shown as a black dot. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

Table V. Mean building height (Z_R), roof width (W_{roof}), road width (W_{road}) and plan area fraction urban (f_{urb}) of the measurement footprint for the three urban sites from the BUBBLE campaign. Corresponding active built (χ_{built})/vegetated (χ_{veg}) indices and UZE computed after Loridan and Grimmond (2011).

Site	f_{urb} (−)	Z_R (m)	W_{roof} (m)	W_{road} (m)	χ_{built} (−)	χ_{veg} (−)	NLCD (1992)	NLCD (2001)	UZE
BSPR	0.84	14.6	11.2	20.2	0.171	0.284	HIR	DHI	1: HD ($\chi_{built} > 0.11$ and $\chi_{veg} < 0.43$)
BSPA	0.69	12.5	14.8	17.2	0.130	0.449	LIR	DMI	2: MD ($\chi_{built} > 0.11$ and $\chi_{veg} > 0.43$)
ALLS	0.47	7.5	11.9	17.5	0.085	0.677	LIR	DLI	3: LD ($\chi_{built} < 0.11$)

repeated with both vegetation classes). It was concluded that 'mixed shrubland-grassland' should be used for the MD and LD classes, otherwise MOSCEM attempts to reduce the positive bias in Q_E (Figure 4) by optimizing to unreasonably large urban fractions (e.g. $f_{urb} > 0.6$ for LD). Changing vegetation class had less impact for HD optimization runs but a better performance ($RMSE_\Sigma$) is obtained with 'cropland/grassland mosaic'. Thus for Stage 4 (and 5) 'mixed shrubland-grassland' is used for LD and MD sites (Table IV) and 'cropland/grassland mosaic' for HD sites.

The use of a distinct vegetation class depending on the urban category is likely to produce strong spatial gradients in the evaporation pattern of adjacent grid cells when

running the scheme 'on-line' with WRF. Associating the HD category with the 'mixed shrubland-grassland' vegetation class (i.e. same as LD/MD) slightly reduces the 'off-line' $RMSE_{\sum}$ performance (not shown) but might be preferred for some on-line applications where these gradients become unrealistic.

3.3.1. *Low density category*

Figure 5(a) shows the first step of the MOSCEM optimization procedure for LD sites with 1000 different combinations of the 19 parameters (parameters 2–20, Table IV) and f_{urb} fixed to site-specific values. The trade-off in simulating Q^* and Q_H results in no combination having both RMSEs at their minimum; for most parameters this trade-off is also obvious from the values selected by the best combinations for Q^* and Q_H (inset, Figure 5(a)).

Here the minimized sum of the RMSE ($RMSE_{\sum}$) is the criterion used to select parameter values, as the relative performance is treated equally. However, a user might select a particular weighted $RMSE_{\sum}$ to prioritize a particular flux(es). The ten combinations with lowest $RMSE_{\sum}$ (top 1%) are shown in purple in Figure 5(a) (and inset), with the best indicated with an asterisk. Using the top 1% combinations, the building height (Z_R, parameter 2 (p 2)) and roof width (W_{roof}, p 4) are consistently kept to very low values for the LD category, while the range of street width (W_{road}, p 3) selected is towards the higher values. The geometry selected by MOSCEM is therefore consistent with *a priori* expectations for that class: small houses with wide streets.

Compared to roof height (3.01 m $< Z_R <$ 3.33 m) and width (5.01 m $< W_{roof} <$ 5.73 m), the range of values selected for the road width (15.02 m $< W_{road} <$ 27.51 m) is considerably wider and is to be linked to the wide range selected for the standard deviation of roof height (σ_Z, p 5). Given the way in which the sensible heat fluxes from the canyon and roof spaces are combined in the SLUCM, both quantities are indirectly connected and any reduction of the canyon roughness via a change in its width is compensated for by an increased roof roughness (larger σ_Z) to provide a similar aggregated value of Q_H. The range of values selected for the roof albedo (α_{roof}, p 6) also contributes to a similar compensation. These highlight the problem of a multi-parameter optimization and suggest the use of the top 1% combinations as preferable when selecting default parameters (rather than only the best option). All combinations within the top 1% perform almost equally (see RMSEs, Figure 5(a)) and therefore any of the ten options is a good candidate to characterize the category at Stage 4. In the particular case of the low-density class, the best option (asterisk, Figure 5(a)) would suggest $\sigma_Z = 3.15$ m and $W_{road} = 15.02$ m; however, one of the other top 1% combinations is preferred ($\sigma_Z = 1.23$ m and $W_{road} = 16.9$ m; see Table IV for the complete list of values) because it is believed to better match the physical properties of the class.

All three albedos (p 6–8) have optimized minimum values that are low, as MOSCEM attempts to compensate for the negative Q^* bias (Figure 3(e)). Similarly, the roof emissivity (ε_{roof}, p 9) is kept smaller. More variability is observed in the road and wall emissivities (p 10, 11) but these do not have a strong impact on the performance of Q^* (Loridan *et al.*, 2010).

For material conductivity (p 12–14) and heat capacity (p 15–17) the optimum roof values are close to their minima whereas road/wall ones are maximized. This was observed also for the single parameter optimization using Marseille data (Loridan *et al.*, 2010). Reduced long-wave emissions from the roof directly reduce Q^* bias, especially at night when turbulence is weak. MOSCEM minimizes both k_{roof} and C_{roof} which reduces the Q^* negative bias (Loridan *et al.*, 2010, see discussion in their section 5.2). However, this causes an increase in daytime Q_H and positive MBE for Stages 4 and 5 (Figure 3(f)). The wall and road surfaces have less direct impact (trapping of radiative fluxes, weaker turbulence inside the canyon), therefore the heat capacity and thermal conductivities maintain higher values which increase the retention of incoming radiative energy without any trade-offs in Q_H model performance (Loridan *et al.*, 2010, their Table V). Material depths (p 18–20) tend to be maximized, retaining more of the incoming radiative energy.

It is important to note that for several parameters the procedure selects values that are very close to the lower/upper limits (e.g. road albedo and roof conductivity; p 8 and p 12). This highlights the role of the preliminary study from Loridan *et al.* (2010) in which an extensive review of typical material properties (e.g. as listed in the ASHRAE (2005) tables) was performed to identify physically reasonable limits. Similarly, both roof height and width have reached what was here set as a lower limit (one storey high and 5 m wide roofs) suggesting again that MOSCEM is trying to compensate for model deficiencies with an extreme choice of parameter values. This underlines the motivation for Stage 5 of this study.

With the 19 parameters set (Table IV), the f_{urb} optimization run is performed with the objective of minimising RMSE for Q^*, Q_H and Q_E. The MOSCEM optimum as f_{urb} increases (0.368 to 0.546) reduces the Q_E RMSE from 39.3 to 29.8 W m^{-2} (Figure 5(b)), as the RMSE for Q_H increases from 38.1 to 42.8 W m^{-2} (trade-off) but the RMSE for Q^* is barely impacted (ΔRMSE < 0.1 W m^{-2}). A decrease in f_{urb} (Figure 5(b)) gives more weight to Noah-simulated fluxes and increases Q_E while reducing Q_H. Here this leads to deterioration in performance for Q_E, suggesting evaporation is overestimated when f_{urb} is less than 0.546. The improvement of Q_H with decreasing f_{urb} also hints at a positive bias in the modelling of Q_H in this range. The $RMSE_{\sum}$ based optimum suggests $f_{urb} = 0.492$ for the LD category (black dot, Figure 5; Table IV).

3.3.2. *Medium and high density categories*

The optimization runs for MD (Figure 6(a)) and HD (Figure 7(a)) categories both have a very steep curve around the lowest RMSE for Q_H, so further improvement in Q_H has a large negative impact on performance of Q^*. Consequently the top 1% combinations identified to minimize $RMSE_{\sum}$ are all located outside of this steep area. Two distinct patterns seem to emerge out of the 1000 optimum solutions (inset, Figures 6(a) and 7(a)): the first are characterized by low, narrow buildings, wide streets (i.e. similar to the LD class, section 3.3.1) and a very high σ_Z (compensating for the low canyon roughness generated by such geometry), while the second, more representative of industrial areas or city centres, has tall wide buildings, narrow streets and a smaller σ_Z. All of the top 1% combinations belong to the second

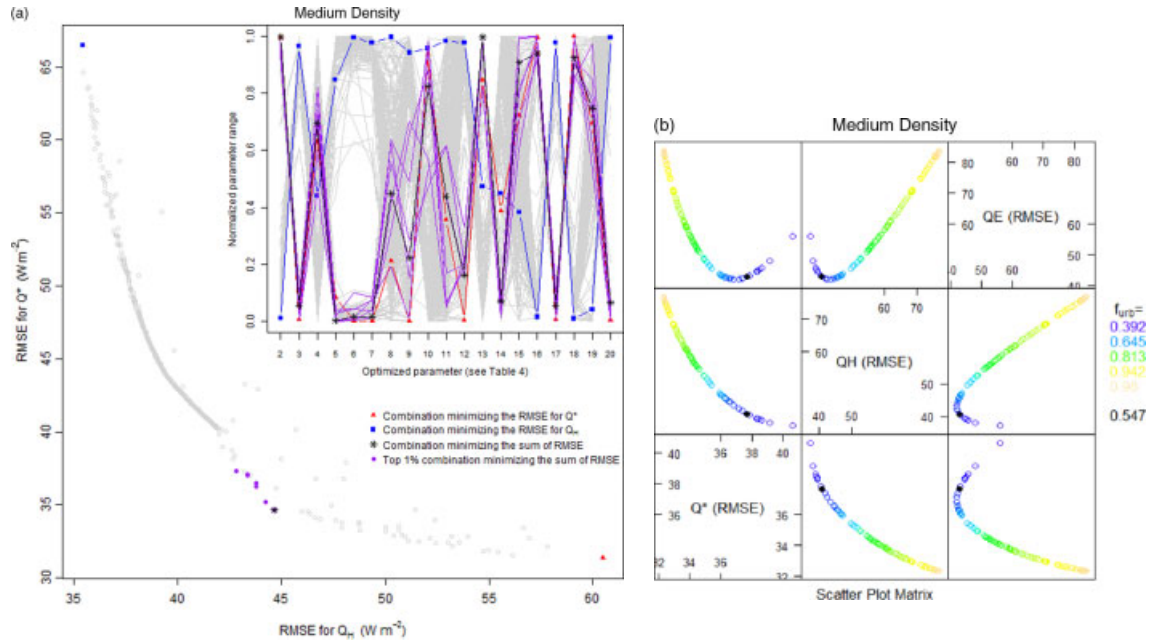

Figure 6. As Figure 5 but for the MD category. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

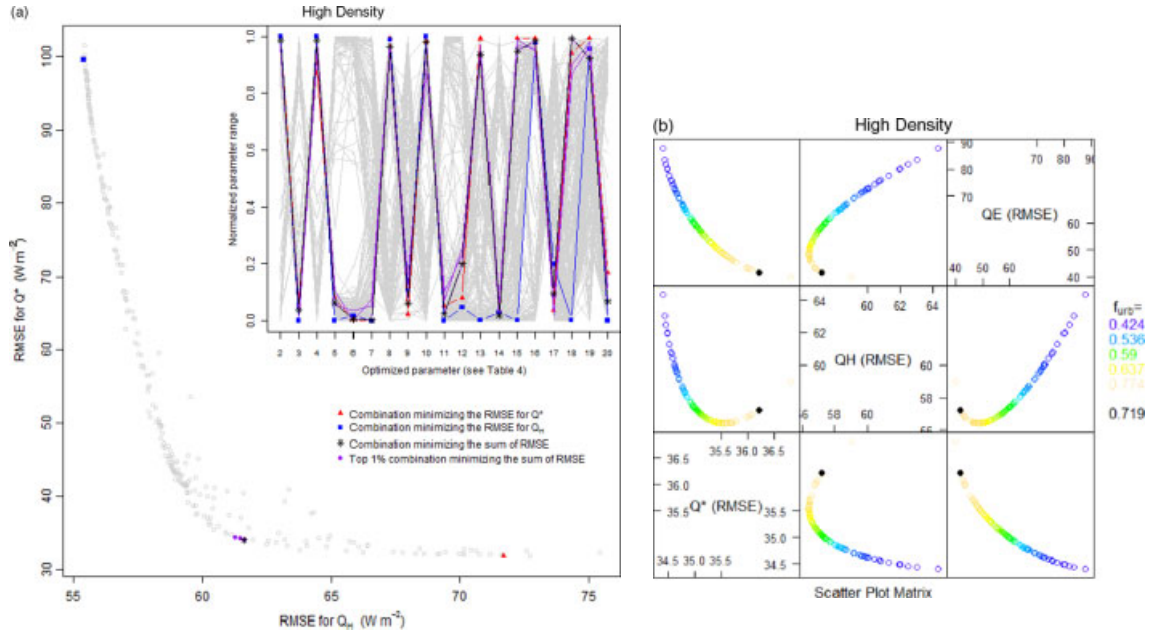

Figure 7. As Figure 5 but for the HD category. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

group. In the MD category some variability is observed in the W_{roof} values ($14.3 < W_{roof} < 21.4$ m) and the emissivity and road albedo values (which have a minimal impact on model performance: Loridan *et al.*, 2010). For all other parameters there is a clustering of the top 1% combinations, which is even more obvious for the HD category where there is a remarkable consistency in the parameter values selected (Figure 7(a), inset) with effectively only two solutions emerging from the optimization, both of which point at a geometry with high and wide buildings and narrow streets. Another interesting feature of the HD optimization is that the combinations minimizing the RMSE for Q^* (red triangle) and Q_H (blue square) only differ significantly in three parameters: heat capacity and thickness of roof materials as well as the thermal conductivity of walls. Yet this creates a drastic change in performance from an RMSE for Q^* of close to 100 W m^{-2} when all three values are set to those that minimize the RMSE for Q_H, to 35 W m^{-2} when using the top 1% combinations (which are close to the configuration minimizing the RMSE for Q^*). This suggests the parameters that characterize the storage capacity of the roof are critical to the partitioning of energy between radiative and turbulent processes.

The triple-objective MD optimization runs show that a decrease in f_{urb} (0.98 to 0.55; i.e. towards the site-specific average f_{urb}) is associated with an improvement in the RMSE for both Q_H and Q_E (better partitioning of turbulent energy). Below $f_{urb} = 0.55$, the RMSE for Q_E begins to increase as the model switches from underestimating to overestimating Q_E. The f_{urb} decrease also triggers a trade-off in Q^* (given the negative bias in Q^*, an increase in f_{urb} is needed to ensure the RMSE decreases) so the resulting minimum RMSE$_\sum$ is when

$f_{urb} = 0.547$. For the HD optimization, an improvement in the model performance for both turbulent fluxes occurs with increasing f_{urb}, toward the site-specific average of 0.8. The threshold when the RMSE for Q_H starts to increase again is $f_{urb} \sim 0.70$ (Figure 7(b)) and the RMSE for Q^* evolves in the opposite direction to the turbulent fluxes. The minimised $RMSE_{\sum}$ is when $f_{urb} = 0.719$.

3.3.3. *Model performance at Stage 4*

The optimized values (Stage 4) show remarkable similarity with the mean f_{urb} values for each class at Stages 2 and 3 (Table IV), especially given the fact that f_{urb} was not constrained ($0 < f_{urb} < 1$ for all three classes) during the procedure. The input class values identified from the minimized $RMSE_{\sum}$ results are consistent with *a priori* expectations in all three cases. For each class the geometry of typical street canyons was well reproduced, when using the top 1% minimized $RMSE_{\sum}$ combinations from the optimization: low buildings/houses with wide streets for LD, and tall buildings with narrow streets for the higher density classes (MD and HD). This shows that Noah/SLUCM responds to changes in observed energy fluxes (magnitude/partitioning) by consistently adapting its surface characteristics to minimize the RMSE statistics.

However, the comparison of the model evaluation statistics for Stage 4i (with site-specific f_{urb}; i.e. after first step of optimization only) and 4 (with optimized class f_{urb} values) again shows that the use of the best information available for f_{urb} does not necessarily lead to the best model performance (Figures 2 and 3). This feature can be attributed to model biases (e.g. lack of Q_E modelling in the urban tile) and is compensated when optimizing f_{urb} for each class (second step of the optimization, Figures 5(b), 6(b) and 7(b); see model performance at Stage 4). Analysis of the individual sites performance (Figure 2) shows that the largest outliers in Stage 4i RMSE are only at four sites, for Q_E: OK03 and AR93/94 which are characterized by the lowest f_{urb} values of the database and VS92 for which, as discussed in section 2.5, measurements were performed under unusually dry conditions.

The model evaluation statistics using all the data for Stage 4 notably show reduced spread (see individual runs, Figure 2; IQR and whiskers, Figure 3). The MOSCEM-optimized Noah/SLUCM using 27 datasets has great consistency (e.g. upper whiskers for the RMSE of the three fluxes all below 60 W m^{-2}, Figure 3). Compared to Stages 1b, 2 and 3, the improvement in the RMSE and MBE for Q^* has been achieved to the detriment of Q_H (increased RMSE and MBE, Figure 3(b) and (f)). The RMSE performance for Q_E is also improved although the MBE is now negative (Figure 3(g)). As expected, Stage 4 provides the best of all results based on $RMSE_{\sum}$ (lowest median and smallest IQR, Figure 3(d)). Biases in the model are, however, evident as the MOSCEM algorithm has an inability to optimize both the radiative and turbulent fluxes.

3.4. *Stage 5: recommendation*

Although Stage 4, by definition, provides the best ability to model Q^*, Q_H and Q_E with the current data (minimized $RMSE_{\sum}$), the parameter values are not always physically realistic for the environments they characterize. Stage 5 reviews the mean values based on observations (Stage 2 and 3, Table IV) in light of the MOSCEM optimization (Stage 4) which accounts for biases in the model. In that sense, parameters which have reached their upper (lower) limit and for which Stages 2/3 would suggest a much lower (higher) value were (arbitrarily) adjusted accordingly. However, consistency is maintained between categories when selecting parameter values. This ensures that even though model deficiencies are being compensated for by the choice of input values, these are still representative of the environment they characterize. The three urban categories (Stage 5b) are characterized by:

1. HD areas have tall buildings, large roof areas and narrow streets ($Z_R = 18$ m; $W_{roof} = 22$ m; $W_{road} = 8$ m). The roof height variability ($\sigma_Z = 3$ m) corresponds to one storey and the vegetated cover is only 25% of the plan area ($f_{urb} = 0.75$) which should be modelled using the 'cropland/grassland mosaic'. The albedo values are low ($\alpha_{roof} = \alpha_{wall} = 0.1$; $\alpha_{road} = 0.15$), which compensates for model biases, while material heat capacities and thermal conductivities are increased compared to their Stage 1b values ($C_{roof} = C_{road} = 1.5 \times 10^6$ J m^{-3} K^{-1}; $C_{wall} = 1.4 \times 10^6$ J m^{-3} K^{-1}; $k_{roof} = k_{road} = 0.8$ W m^{-1} K^{-1}; $k_{wall} = 1.0$ W m^{-1} K^{-1}). Note that road properties are kept constant for all three classes.
2. MD areas have street canyon geometry that is similar to the HD class ($Z_R = 15$ m; $W_{roof} = 20$ m; $W_{road} = 10$ m) but a greater vegetation fraction ($f_{urb} = 0.60$) and more consistency in roof heights ($\sigma_Z = 1.5$ m). The recommended vegetation class is the 'mixed shrubland-grassland'. Although the albedo and emissivity values are identical to HD, the thermal conductivities and heat capacities are lower for roof and wall materials ($C_{roof} = C_{wall} = 1.2 \times 10^6$ J m^{-3} K^{-1}; $k_{roof} = 0.4$ W m^{-1} K^{-1}; $k_{wall} = 1.0$ W m^{-1} K^{-1}). This is the default class used for all sites in Stage 5a.
3. LD areas are more characteristic of residential areas, i.e. low buildings, with reduced roof coverage, wide roads and even lower variability in roof heights ($Z_R = 6$ m; $W_{roof} = 8$ m; $W_{road} = 15$ m; $\sigma_Z = 1$ m). Vegetation covers 50% of the plan area ($f_{urb} = 0.50$) which is very consistent with the mean values from Stages 1b, 2 and 3 (Table IV). Vegetation class for Noah is the 'mixed shrubland-grassland'. The albedo of roof surfaces is larger than for the other two categories ($\alpha_{roof} = 0.15$) but the heat capacities are lower ($C_{roof} = 1.0 \times 10^6$ J m^{-3} K^{-1}) as suggested by the Stage 4 results.

When these parameter values are used with all the data, there is a minimal deterioration from Stage 4 with Q_E the least impacted (Stage 5b, Figures 2 and 3). The Q^* and Q_E RMSE performances are both improved from the Stages 1b, 2 and 3 results (lower median and IQR, Figure 3(a) and (c)), whereas Q_H is slightly poorer (Figure 3(b)). The $RMSE_{\sum}$ suggests Stage 5b to be the best performing (after Stage 4) with a lower median (and upper quartile) than earlier stages. As for Stages 0 and 1, the impact of having only one class (Stage 5a) is assessed and results show a great improvement when refining the classes (5a to 5b, Figure 3). As with previous comments about Stages 0c, 1c and 4i, the use of site-specific f_{urb} values (Stage 5c) leads to performances

slightly poorer than with optimum class values (Stage 5b; Figure 3(d)). Yet statistics from Stage 5c are very close to Stage 4i, again showing that the parameter values selected for Stage 5 do not strongly deteriorate model performance. It also shows that a user willing to provide site-specific f_{urb} to ensure a more realistic surface representation (e.g. when looking at landscape adaptation measures) could do so knowing that the model performance stays comparable to Stage 5b. As for Stage 4i (see section 3.3.3) analysis of Figure 2 reveals that a significant increase in RMSE only occurs for Q_E and the sites of OK03, AR93/94 and VS92. Given the good performance and the wide range of sites considered, it is recommended that the number of classes in WRF is kept at three but the default values are changed to the Stage 5b values. It should be noted that given the lack of observations in very dense urban areas (i.e. those classified as TBD by Jackson *et al.* (2010)) the HD class might not adequately represent very dense city centres like Manhattan, Hong Kong or Tokyo. As highlighted by Jackson *et al.* (2010) there are relatively few such environments covering an area of at least 1 km^2 and these would therefore only represent a few grid boxes at current NWP resolutions. However, WRF users are advised to refine SLUCM parameters for those grid cells.

The variety of locations, morphology and conditions available from the current dataset provides a robust assessment of the scheme's ability to cope with the range of urban environments involved in simulations using the WRF model. Ordered in terms of their RMSE at Stage 5b, the histograms of Figure 2 give an overview of the scheme's ability to simulate the key fluxes required as lower boundary conditions in an atmospheric model such as WRF, and show that although some variability in performance can be observed, the entire range of sites presented here are simulated with RMSEs below ~60 W m^{-2} for Q^*, ~70 W m^{-2} for Q_E, and ~85 W m^{-2} for Q_H. The RMSE upper quartile (Figure 3) is ~40 W m^{-2} for Q^* and Q_E, while it is ~50 W m^{-2} for Q_H. The RMSE medians for the three fluxes are at 40 W m^{-2} (Q_H) or lower (Q^* and Q_E). Such performances are in the range of the PILPS-urban model cohort (Grimmond *et al.*, 2010, 2011) and other off-line model evaluation exercises (Grimmond and Oke, 2002; Masson *et al.*, 2002; Dupont and Mestayer, 2006; Hamdi and Schayes, 2007; Kawai *et al.*, 2009; Porson *et al.*, 2010). This indicates that the SLUCM is performing as well as any state-of-the-art model and suggests that we should expect state-of-the-art performance from the WRF-SLUCM.

It should, however, be noted that the performances at Stage 4/5 (Figure 3) are consistent with known limitations of the scheme. In particular the Q_E negative/Q_H positive biases (Figure 3(f) and (g)) are understandable given the lack of evaporation in the urban tile. If improvement were to be made to the SLUCM these should improve the performance with Stage 5 parameters consistently (i.e. energy transferred from Q_H to Q_E in the urban tile, leading to a reduction of both biases). Similarly the current representation of heat conduction through the different layers of materials composing the urban surface is rather simple (no distinction between layers in terms of heat capacity and conductivities) and likely responsible for the misrepresentation of Q^*. Again, the model biases should respond positively to a modification of the scheme.

Figure 8 groups the model RMSEs and MBEs at Stage 5b by class (Table II) with only one year of data used for Helsinki and Łódź (HE09 and ŁÓ02) to ensure a similar number of data points in each class. The summary statistics ($RMSE_\Sigma$ and MBE_Σ, Figure 8(d) and (h)) are better for the two extreme classes (HD and LD) than for the transitional

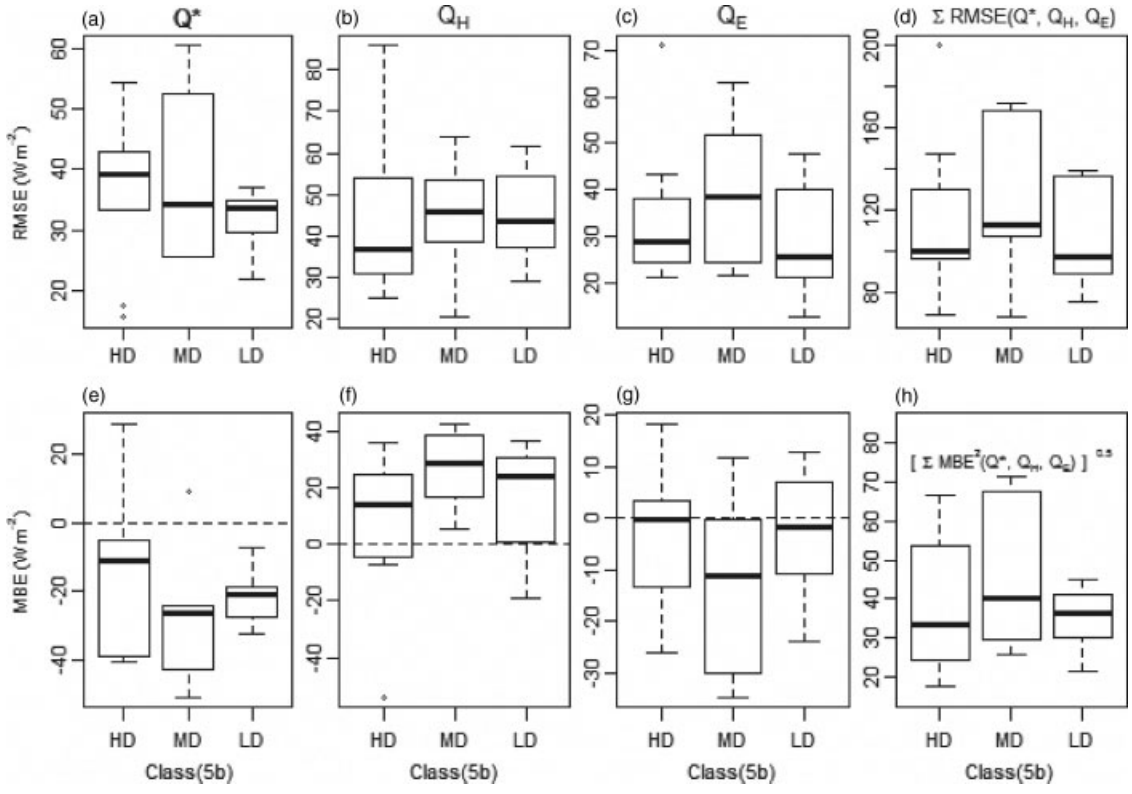

Figure 8. Tukey's (1977) schematic plots summarizing results obtained with Stage 5b parameter values for each class (HD, MD and LD): (a)–(c) RMSE for Q^*, Q_H and Q_E; (d) $RMSE_\Sigma = \Sigma\ RMSE(Q^*, Q_H, Q_E)$; (e)–(g) MBE for Q^*, Q_H and Q_E, and (h) $MBE_\Sigma = \{\Sigma\ MBE^2(Q^*, Q_H, Q_E)\}^{0.5}$.

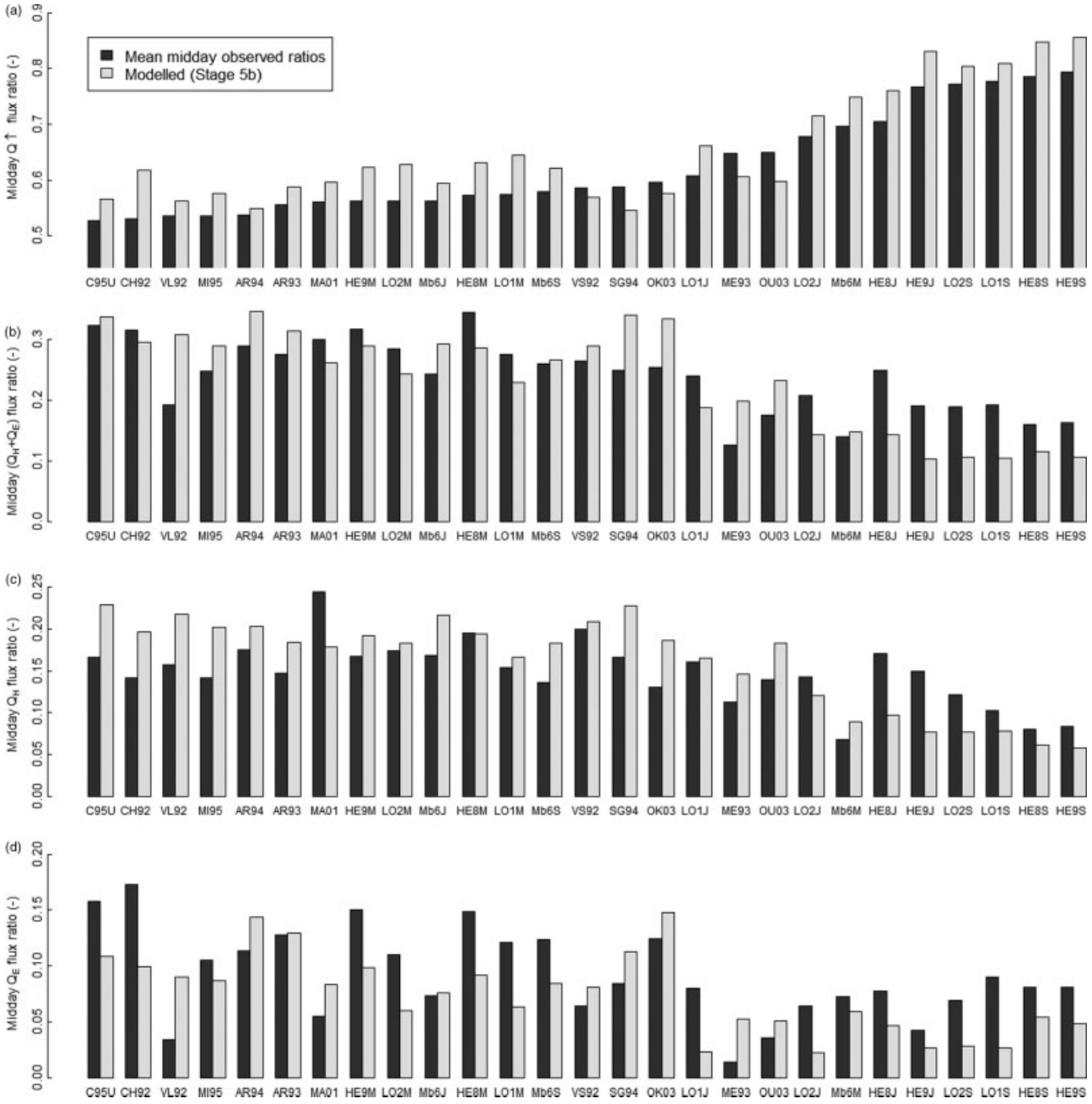

Figure 9. Mean midday (±3 hours around solar noon) simulated (Stage 5b) and observed (a) $Q\uparrow/Q\downarrow = (K\uparrow + L\uparrow)/(K\downarrow + L\downarrow)$, (b) $(Q_H + Q_E)/Q\downarrow$, (c) $Q_H/Q\downarrow$ and (d) $Q_E/Q\downarrow$ flux ratios. Sites (Table II) are ordered by increasing $Q\uparrow/Q\downarrow$ value.

one (MD). This further supports the idea that a refined classification (a-b sub-stage) is beneficial. The HD and LD MBE values for Q_E (Figure 8(g)) suggest the change of vegetation class for Noah is useful in dealing with changes in evaporation regime. Finally, the lack of any strong bias in performance towards one type of environment underlines the ability of the Noah/SLUCM to cope with the range of urban forms.

Given the need for models to perform well across the range of seasons, it is important to remember the RMSEs (Figures 2 and 3) are not independent of the magnitude of the fluxes. The RMSEs are generally higher for spring/summer months (Figure 2) as a direct consequence of larger heat fluxes (Figure 2(a)). The larger incoming radiative forcing, longer day length and the active contribution of vegetation alter the flux magnitudes. Following Loridan and Grimmond (2011), the fluxes are normalized by the total incoming radiative energy $Q\downarrow = K\downarrow + L\downarrow$ so that the sites and seasons can be compared independently of the amount of energy they receive. By ranking the 27 datasets by increasing mean midday (±3 hours around solar noon) observed $Q\uparrow/Q\downarrow$ ratios ($Q\uparrow = K\uparrow + L\uparrow$), the autumn/winter datasets (Figure 9, from ŁÓ1J to HE9S) are clearly distinguished from the spring/summer ones with higher ratios. The mean daytime portion of observed incoming energy dissipated as $Q\uparrow$ (Figure 9(a)) is far greater in the autumn/winter months ($\sim$0.6–0.8$Q\downarrow$) and therefore the turbulent flux ratios (Figure 9(b)–(d)) are correspondingly reduced compared to the spring/summer data. Deviation from this general trend is most likely due to restricted water availability (e.g. dry conditions in VL92, VS92, ME93; Figure 9(d)), different storage capacities (e.g. ME93) and/or mis-estimation of the anthropogenic heat contribution to the fluxes (e.g. HE9J; Figure 9(c)). The

Table VI. RMSE statistics ($Q^*/Q_H/Q_E/RMSE_\sum$) when running the Noah/SLUCM at Stage 5b for the three BUBBLE sites (Table V) and when using each of the three UZEs.

RMSE (W m^{-2})	LD				MD				HD			
	Q^*	Q_H	Q_E	$RMSE_\sum$	Q^*	Q_H	Q_E	$RMSE_\sum$	Q^*	Q_H	Q_E	$RMSE_\sum$
ALLS	**32.5**	**24.7**	**26.0**	**83.2**	29.8	30.6	29.3	89.7	32.2	29.7	38.0	99.9
BSPA	48.5	56.0	58.4	162.9	**43.6**	**44.7**	**50.6**	**138.9**	46.0	46.9	49.0	141.9
BSPR	42.2	52.2	40.6	135.0	37.8	40.5	30.8	109.1	**35.5**	**36.3**	**29.7**	**101.5**
Mean	41.1	44.3	41.7	127.1	**37.0**	**38.6**	**36.9**	**112.5**	37.9	37.6	38.9	114.4

The mean statistics per UZE are also reported. Best $RMSE_\sum$ is shown in bold for each row.

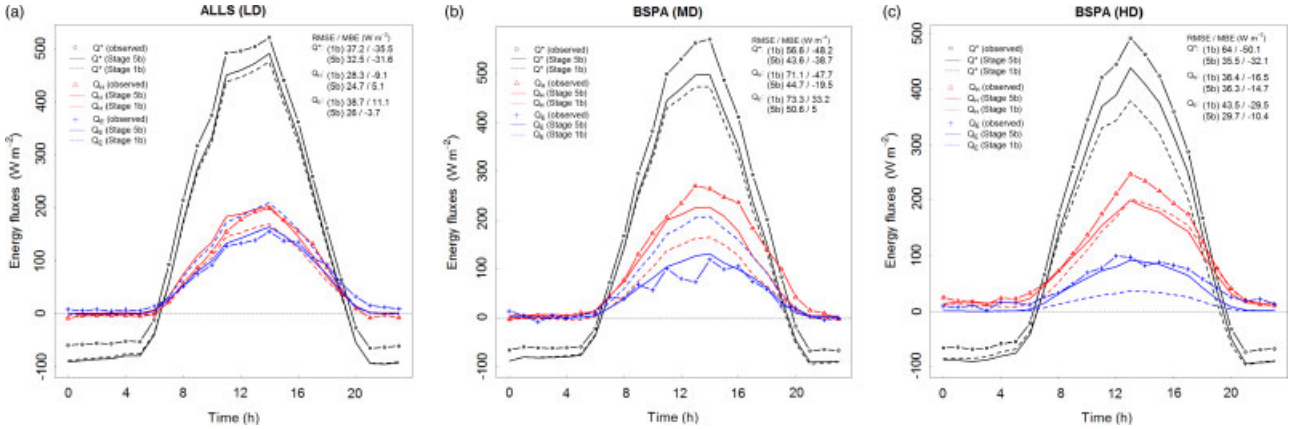

Figure 10. Observed and Noah/SLUCM simulated (Stage 1b and 5b) mean diurnal Q^*, Q_H and Q_E fluxes at the three urban BUBBLE sites during the June–July 2002 IOP (Christen and Vogt, 2004). RMSE/MBE model performance statistics are reported. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

features of the observed fluxes are discussed elsewhere (Loridan and Grimmond, 2011); here we are concerned with the ability of the Noah/SLUCM to reproduce this overall trend. Although errors exist in the simulated flux ratios at each site, the Noah/SLUCM captures the general seasonal trend remarkably well: both the increase in $Q\uparrow/Q\downarrow$ (Figure 9(a)) and the decrease in $(Q_H + Q_E)/Q\downarrow$ ratios (Figure 9(b)) for the autumn/winter months are very clear. Given the parametrization of seasonal vegetation evolution (Noah) and explicit representation of shading patterns in street canyons (SLUCM), the scheme is able to cope with seasonal changes in energy flux partitioning. This provides positive support to the modelling across the range of seasons that are required for applications such as NWP. Analysis of the individual datasets shows the scheme generally overestimates the mean midday $Q\uparrow/Q\downarrow$ (Figure 9(a)), overestimates the Q_H contribution in spring/summer, and generally underestimates it in the autumn/winter (Figure 9(c)). For Q_E no clear trend is apparent in spring/summer when biases would be likely attributable to local water availability, while autumn/winter datasets suggest a model underestimation (Figure 9(d)).

4. Independent evaluation

Three urban sites from the Basel Urban Boundary Layer Experiment (BUBBLE: Christen and Vogt, 2004; Rotach *et al.*, 2004) are used (Table V, section 2.7) to independently assess the recommendations made here.

The mean diurnal evolution of simulated fluxes from Stages 1b and 5b are compared to observations collected during the IOP (Figure 10). The general diurnal evolution/magnitude of the three fluxes for all three sites at Stage 5b are predicted well, with RMSE statistics all between ~25 and 50 W m^{-2}. This shows the ability of the scheme with Stage 5b parameters to cope with the three types of environments and represents a clear improvement on the Stage 1b runs with reduced RMSE and MBE statistics for all fluxes and sites. The underestimation of Q^*, most obvious for the two most urban sites (BSPR and BSPA), is noticeable in the daytime maximum and night-time minimum values; yet it is considerably reduced when compared to Stage 1b runs, highlighting a clear improvement from the proposed parameter values. The Q^* bias has a direct impact on modelled Q_H flux (daytime underestimation) at BSPR and BSPA while the suburban Q_H magnitude is better predicted (ALLS). The correct estimation of the daytime maximum Q_E in all three cases also highlights the capacity of the scheme, with the f_{urb} values from Stage 5b, to simulate distinct energy partitioning patterns.

To further test the methodology (UZE and Stage 5b parameters) each of the BUBBLE sites was run with the three UZEs. The RMSE for the three fluxes and $RMSE_\sum$ are reported in Table VI. For all fluxes and all sites apart from Q^* in ALLS and Q_E in BSPA, the RMSE statistics suggest the sites have been assigned to the most appropriate class. This feature is particularly clear when considering the $RMSE_\sum$ since the minimum value is always reached for the UZE assigned. Also of interest are the mean RMSE values across the three sites as they confirm the MD class as the best default to use when no refinement is done (i.e. sub-stages a).

5. Conclusion

An extensive off-line evaluation of the Noah/SLUCM using 27 datasets representative of a range of urban environments, climatic conditions and geographical locations was conducted. The five-stage procedure was designed to evaluate the impact of increasing complexity when deriving input parameter values (i.e. default set of values for either one or three urban classes; region-dependent classes from the Jackson *et al.* (2010) database; site-specific; or optimized using the MOSCEM algorithm). From this we conclude:

- Although a clear trade-off occurs between the modelling of Q^* and Q_H (Figure 3(a), (b), (e) and (f)), the overall performance of the scheme (modified version of the model, Stage 1) is improved compared to the original (Stage 0) form (Figure 3(d)).
- Performance is greatly improved by using three classes of urban form (refined sub-stages 0b, 1b and 5b) over using one class (default sub-stages 0a, 1a and 5a). In particular the MBE improvement for both Q_H and Q_E suggests better partitioning of the turbulent energy release. This is attributable to the refined urban fraction (f_{urb}) value which is the major change between class parameter values (see Stages 0, 1; Tables III and IV). This conclusion agrees with Grimmond *et al.* (2010, 2011) who highlighted the critical implications of the f_{urb} estimation. The use of site-specific urban fractions (sub-stages c) do not, however, provide further improvements, suggesting the scheme is not able to fully benefit from the best information available.
- Similarly, more site-specific parameter estimations (Stages 2 and 3) are associated with relatively low RMSE reductions compared to the initial sub-stage refinement (a to b) leading to very similar performances for 1b, 1c, 2 and 3. This suggests that improvements beyond the initial refinement of classes come at the price of intensive effort for parameter estimation (e.g. from observation or geographic information systems (GIS) analysis) and, depending on the application, may not be justified given model biases.
- Even after using an optimization algorithm (MOSCEM) to minimize the RMSE of Q^*, Q_H and Q_E (multi-parameter/multi-objective optimization), the model performance fails to increase much further. Analysis of the optimum solutions demonstrate trade-offs between Q^* and Q_H which confirms the single parameter optimization results with the Marseille data (Loridan *et al.*, 2010).
- Additional trade-off issues arise when trying to identify an optimum f_{urb} value; these are intrinsic to the tile approach employed in the Noah/SLUCM.
- A systematic tendency of the scheme to underestimate Q^* is observed through Stages 1–5 and is only compensated in Stage 0 via a strong positive Q_H bias (trade-off). This tendency is also obvious from the choice of values arising from the MOSCEM optimization with many parameters set to their limits to compensate for this model deficiency (e.g. roof albedo values).

In Stage 5b a set of recommended parameter values (Table IV) were identified for three categories of urban areas (UZE). The performance of Noah/SLUCM using these is in line with Stages 1b–4 (Figure 3) with RMSE medians across the 27 datasets for the three fluxes at 40 W m^{-2} (Q_H) or lower (Q^* and Q_E). This compares well with the performance reported for similar schemes (Grimmond *et al.*, 2010, 2011). A methodology was evaluated for objectively assigning a site to a category (i.e. through estimation of its active surface indices (Loridan and Grimmond, 2011)) which is considered a great advantage for use at new sites (off-line) or grid cells from an NWP domain (on-line). The independent evaluation of the performance of Noah/SLUCM with Stage 5b parameters (section 4) further supports the recommendations that (1) three classes are appropriate for characterizing the urban environment, and (2) that the parameter values given in Table IV (Stage 5b) should be adopted as default values in WRF. They are also shown to reproduce energy flux ratios including seasonal variations in flux partitioning (Figure 9).

With the Table IV values included in WRF, WRF/SLUCM users could then decide which UZE a particular grid cell should be assigned, from consideration of either (1) its typical physical characteristics (e.g. urban morphology, surface area covered by vegetation), or (2) computation of the active surface indices (FRAISE scheme: Loridan and Grimmond, 2011) if further information about the surface is available (e.g. from a GIS database or observations). As a result, high-resolution urban WRF simulations using the Noah/SLUCM should benefit from a better estimation of intra-urban heterogeneities in flux partitioning which can have major impact on, for instance, the structure of the modelled boundary layer.

Acknowledgements

We thank all those involved in the collection and processing of the flux data (in particular Brian Offerle, Krzysztof Fortuniak, Andy Coutts, Jason Berringer, Leena Järvi, Annika Nordbo and Timo Vesala; and Andreas Christen and Roland Vogt), Johannes Feddema for the release of the Jackson *et al.* (2010) database and both Jasper Vrugt and Luis Bastidas for the development and sharing of the MOSCEM software. Thanks to Fei Chen, Mukul Tewari, Kevin Manning and Susanne Grossman-Clarke for useful discussions about WRF implementation. We thank a reviewer for a very detailed review and discussion. Financial support for this project was provided by the US National Science Foundation (ATM-0710631) the Met Office (P001550) and EU FP7-ENV-2007-1 BRIDGE (211345).

An R version of the FRAISE scheme is available at: http://geography.kcl.ac.uk/micromet/index.htm.

References

Allen L, Lindberg F, Grimmond CSB. 2011. Global to city scale urban anthropogenic heat flux: Model and variability. *Int. J. Climatol.* **31**: 1990–2005, DOI: 10.1002/joc.2210.

Anderson V. 2009. '*The exacerbation of anthropogenic heat fluxes as a function of interior climate controls and building envelope*.' Master's thesis, M. S. Geography, Atmospheric Science Program, Indiana University, Bloomington. Obtainable from Sue.Grimmond@kcl.ac.uk

ASHRAE. 2005. *ASHRAE Handbook: 2005 Fundamentals*. Am. Soc. of Heating, Refrigerating and Air conditioning Engineers.

Chen F, Avissar R. 1994a. The impact of land-surface wetness heterogeneity on mesoscale heat fluxes. *J. Appl. Meteorol.* **33**: 1323–1340.

Chen F, Avissar R. 1994b. Impact of land-surface moisture variability on local shallow convective cumulus and precipitation in large-scale models. *J. Appl. Meteorol.* **33**: 1382–1401.
Chen F, Dudhia J. 2001. Coupling an advanced land surface–hydrology model with the Penn State-NCAR MM5 modelling system. Part 1: Model implementation and sensitivity. *Mon. Weather Rev.* **129**: 569–585.
Chen F, Mitchell K, Schaake J, Xue YK, Pan H-L, Koren V, Duan QY, Ek M, Betts A. 1996. Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.* **101**: 7251–7268.
Chen F, Kusaka H, Bornstein R, Ching J, Grimmond CSB, Grossman-Clarke S, Loridan T, Manning KW, Martilli A, Miao SG, Sailor D, Salamanca FP, Taha H, Tewari M, Wang XM, Wyszogrodzki AA, Zhang CL. 2011. The integrated WRF/urban modelling system: Development, evaluation, and applications to urban environmental problems. *Int. J. Climatol.* **31**: 273–288.
Christen A, Vogt R. 2004. Energy and radiation balance of a central European city. *Int. J. Climatol.* **24**: 1395–1421.
Coutts AM, Beringer J, Tapper NJ. 2007a. Characteristics influencing the variability of urban CO_2 fluxes in Melbourne, Australia. *Atmos. Environ.* **41**: 51–62.
Coutts AM, Beringer J, Tapper NJ. 2007b. Impact of increasing urban density on local climate: Spatial and temporal variations in the surface energy balance in Melbourne, Australia. *J. Appl. Meteorol. Clim.* **46**: 477–493.
Deardorff JW. 1978. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *J. Geophys. Res.* **83**(C4): 1889–1903.
Dupont S, Mestayer PG. 2006. Parameterization of the urban energy budget with the submesoscale soil model. *J. Appl. Meteorol. Clim.* **45**: 1744–1765.
Ellefsen R. 1990. Mapping and measuring buildings in the canopy boundary layer in ten US cities. *Energ. Buildings* **16**: 1025–1049.
Grimmond CSB, Oke TR. 1995. Comparison of heat fluxes from summertime observations in the suburbs of four North American cities. *J. Appl. Meteorol.* **34**: 873–889.
Grimmond CSB, Oke TR. 1999. Aerodynamic properties of urban areas derived from analysis of surface form. *J. Appl. Meteorol.* **38**: 1262–1292.
Grimmond CSB, Oke TR. 2002. Turbulent heat fluxes in urban areas: Observations and a Local-scale Urban Meteorological Parameterization Scheme (LUMPS). *J. Appl. Meteorol.* **41**: 792–810.
Grimmond CSB, Oke TR, Cleugh HA. 1993. The role of 'rural' in comparisons of observed suburban–rural flux differences. *Int. Assoc. Hydrolog. Sci. Publ.* **212**: 165–174.
Grimmond CSB, Souch C, Grant R, Heisler G. 1994. '*Local scale energy and water exchanges in a Chicago neighbourhood*.' Ch 4, pp 41–61 in *Chicago's Urban Forest Ecosystem: Results of the Chicago Urban Forest Climate Project*, McPherson GE, Nowak DJ, Rowntree RA (eds). USDA, Forest Service, Northeastern Forest Experiment Station: Gen. Tech. Rep. NE-186. Available online at http://www.treesearch.fs.fed.us/pubs/4285
Grimmond CSB, Souch C, Hubble M. 1996. Influence of tree cover on summertime surface energy balance fluxes, San Gabriel Valley, Los Angeles. *Clim. Res.* **6**: 45–57.
Grimmond CSB, Salmond JA, Oke TR, Offerle B, Lemonsu A. 2004a. Flux and turbulence measurements at a densely built-up site in Marseille: Heat, mass (water and carbon dioxide), and momentum. *J. Geophys. Res.* **109**: D24101, DOI: 10.1029/2004JD004936.
Grimmond CSB, Su H-B, Offerle B, Crawford B, Scott S, Zhong S, Clements C. 2004b. 'Variability of sensible heat fluxes in a suburban area of Oklahoma City.' Paper J7.2 in *Symp. on Planning, nowcasting and forecasting in the urban zone, Eighth Symp. on Integrated observing and assimilation systems for atmosphere, oceans, and land surface*. Am. Meteorol. Soc: Boston, MA.
Grimmond CSB, Blackett M, Best MJ, Barlow J, Baik J-J, Belcher SE, Bohnenstengel SI, Calmet I, Chen F, Dandou A, Fortuniak K, Gouvea ML, Hamdi R, Hendry M, Kawai T, Kawamoto Y, Kondo H, Krayenhoff ES, Lee S-H, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Pigeon G, Porson A, Ryu Y-H, Salamanca F, Shashua-Bar L, Steeneveld G-J, Tombrou M, Voogt J, Young D, Zhang N. 2010. The International Urban Energy Balance Models Comparison Project: First results from Phase 1. *J. Appl. Meteorol. Clim.* **49**: 1268–1292.
Grimmond CSB, Blackett M, Best MJ, Baik J-J, Belcher SE, Beringer J, Bohnenstengel SI, Calmet I, Chen F, Coutts AM, Dandou A, Fortuniak K, Gouvea ML, Hamdi R, Hendry M, Kanda M, Kawai T, Kawamoto Y, Kondo H, Krayenhoff ES, Lee S-H, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Ooka R, Pigeon G, Porson A, Ryu Y-H, Salamanca F, Steeneveld G-J, Tombrou M, Voogt JA, Young DT, Zhang N. 2011. Initial results from Phase 2 of the international urban energy balance model comparison. *Int. J. Climatol.* **31**: 244–272.
Hamdi R, Schayes G. 2007. Validation of Martilli's urban boundary layer scheme with measurements from two mid-latitude European cities. *Atmos. Chem. Phys.* **7**: 4513–4526.
Jackson TL, Feddema JJ, Oleson KW, Bonan GB, Bauer JT. 2010. Parameterization of urban characteristics for global climate modeling. *Ann. Assoc. Am. Geogr.* **100**: 848–865.
Kanda M, Kanega M, Kawai T, Moriwaki R, Sugawara H. 2007. Roughness length for momentum and heat derived from outdoor urban scale models. *J. Appl. Meteorol. Clim.* **46**: 1067–1079.
Kawai T, Ridwan MK, Kanda M. 2009. Evaluation of the simple urban energy balance model using selected data from 1-yr flux observations at two cities. *J. Appl. Meteorol. Clim.* **48**: 693–715.
King T, Grimmond CSB. 1997. 'Transfer mechanisms over an urban surface for water vapor, sensible heat, and momentum.' Pp 455–456 in *Preprints, 12th Conf. on Boundary layers and turbulence, Vancouver, BC, Canada*. Am. Meteorol. Soc: Boston, MA.
Kusaka H, Kimura F. 2004. Thermal effects of urban canyon structure on the nocturnal heat island: Numerical experiment using a mesoscale model coupled with an urban canopy model. *J. Appl. Meteorol.* **43**: 1899–1910.
Kusaka H, Kondo H, Kikegawa Y, Kimura F. 2001. A simple single-layer urban canopy model for atmospheric models: Comparison with multi-layer and slab models. *Boundary-Layer Meteorol.* **101**: 329–358.
Lemonsu A, Grimmond CSB, Masson V. 2004. Modeling the surface energy balance of the core of an old Mediterranean city: Marseille. *J. Appl. Meteorol.* **43**: 312–327.
Loridan T, Grimmond CSB. 2011. Characterization of energy flux partitioning in urban environments: Links with surface seasonal properties. *J. Appl. Meteorol. Clim.*, DOI: 10.1175/JAMC-D-11-038.1.
Loridan T, Grimmond CSB, Grossman-Clarke S, Chen F, Tewari M, Manning K, Martilli A, Kusaka H, Best MJ. 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. *Q. J. R. Meteorol. Soc.* **136**: 997–1019.
Loridan T, Grimmond CSB, Offerle BD, Young DT, Smith TEL, Järvi L, Lindberg F. 2011. Local-scale Urban Meteorological Parameterization Scheme (LUMPS): Longwave radiation parameterization and seasonality-related developments. *J. Appl. Meteorol. Clim.* **50**: 185–202.
Macdonald RW, Griffiths RF, Hall DJ. 1998. An improved method for the estimation of surface roughness of obstacle arrays. *Atmos. Environ.* **32**: 1857–1864.
Mahfouf J-F, Richard E, Mascart P. 1987. The influence of soil and vegetation on the development of mesoscale circulations. *J. Clim. Appl. Meteorol.* **26**: 1483–1495.
Masson V. 2000. A physically-based scheme for the urban energy budget in atmospheric models. *Boundary-Layer Meteorol.* **94**: 357–397.
Masson V, Grimmond CSB, Oke TR. 2002. Evaluation of the Town Energy Balance (TEB) scheme with direct measurements from dry districts in two cities. *J. Appl. Meteorol.* **41**: 1011–1026.
Miao SG, Chen F, LeMone MA, Tewari M, Li QC, Wang YC. 2009. An observational and modeling study of characteristics of urban heat island and boundary layer structures in Beijing. *J. Appl. Meteorol. Clim.* **48**: 484–501.
National Climatic Data Center (NCDC). 2009. http://www.ncdc.noaa.gov/oa/ncdc.html (Accessed 24 August 2009).
Newton T, Oke TR, Grimmond CSB, Roth M. 2007. The suburban energy balance in Miami, Florida. *Geogr. Ann.* **89A**: 331–347.
NLCD. 1992. http://www.epa.gov/mrlc/definitions.html (Accessed December 2010). NLCD. 2001. http://www.epa.gov/mrlc/definitions.html (Accessed December 2010).
Noilhan J, Planton S. 1989. A simple parameterization of land surface processes for meteorological models. *Mon. Weather Rev.* **117**: 536–549.
Offerle B. 2003. '*The energy balance of an urban area: Examining temporal and spatial variability through measurements, remote sensing and modeling*.' PhD thesis, Dept. of Geography, Indiana University. Obtainable from Sue.Grimmond@kcl.ac.uk
Offerle B, Jonsson P, Eliasson I, Grimmond CSB. 2005. Urban modification of the surface energy balance in the West African Sahel: Ouagadougou, Burkina Faso. *J. Climate* **18**: 3983–3995.
Offerle B, Grimmond CSB, Fortuniak K, Klysik K, Oke TR. 2006. Temporal variations in heat fluxes over a central European city centre. *Theor. Appl. Climatol.* **84**: 103–115.

Oke TR. 2004. '*Initial guidance to obtain representative meteorological observations at urban sites*.' IOM Report 81, World Meteorological Organization: Geneva.

Oke TR, Spronken-Smith R, Jauregui E, Grimmond CSB. 1999. Recent energy balance observations in Mexico City. *Atmos. Environ.* **33**: 3919–3930.

Oleson KW, Bonan GB, Feddema J, Vertenstein M, Grimmond CSB. 2008. An urban parameterization for a global climate model. Part 1: Formulation and evaluation for two cities. *J. Appl. Meteorol. Clim.* **47**: 1038–1060.

Porson A, Clark PA, Harman IN, Best MJ, Belcher SE. 2010. Implementation of a new urban energy budget scheme into MetUM. Part II: Validation against observations and model intercomparison. *Q. J. R. Meteorol. Soc.* **136**: 1530–1542.

Rotach MW, Gryning S-E, Batchvarova E, Christen A, Vogt R. 2004. Pollutant dispersion close to an urban surface – the BUBBLE tracer experiment. *Meteorol. Atmos. Phys.* **87**: 39–56.

Sellers PJ, Mintz Y, Sud YC, Dalcher A. 1986. A simple biosphere model (SIB) for use within general circulation models. *J. Atmos. Sci.* **43**: 505–531.

Sellers WD. 1972. *Physical Climatology*. The University of Chicago Press: Chicago, IL.

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, Huang XY, Wang W, Powers JG. 2008. '*A description of the advanced research WRF version 3*.' National Center for Atmospheric Research, Tech. Note NCAR/TN-475+STR, 113 pp.

Stewart ID, Oke TR. 2009. 'Newly developed "thermal climate zones" for defining and measuring urban heat island magnitude in the canopy layer.' Paper J8.2A in *Preprints, 8th Am. Meteorol. Soc. Symposium on the urban environment, Phoenix, Arizona, USA*. http://ams.confex.com/ams/pdfpapers/150476.pdf

Tewari M, Chen F, Kusaka H. 2006. 'Implementation and evaluation of a single-layer urban canopy model in WRF/Noah.' *WRF Users Workshop 2006*.

Tukey JW. 1977. *Exploratory Data Analysis*. Addison-Wesley: Reading, MA.

Vesala T, Järvi L, Launiainen S, Sogachev A, Rannik Ü, Mammarella I, Siivola E, Keronen P, Rinne J, Riikonen A, Nikinmaa E. 2008. Surface–atmosphere interactions over complex urban terrain in Helsinki, Finland. *Tellus* **60B**: 188–199.

Vrugt JA, Gupta HV, Bastidas LA, Bouten W, Sorooshian S. 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resour. Res.* **39**(12): 14, DOI: 10.1029/2002WR001746.

Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences: An introduction*. Academic Press: London.

