



Urban signals in high-resolution weather and climate simulations: role of urban land-surface characterisation

Denise Hertwig¹ · Sue Grimmond¹ · Margaret A. Hendry² · Beth Saunders¹ · Zhengda Wang^{1,3} · Marine Jeoffrion^{1,4} · Pier Luigi Vidale¹ · Patrick C. McGuire^{1,5,6} · Sylvia I. Bohnenstengel⁷ · Helen C. Ward⁸ · Simone Kotthaus⁹

Received: 27 November 2019 / Accepted: 8 June 2020 / Published online: 7 August 2020
© The Author(s) 2020

Abstract

Two urban schemes within the Joint UK Land Environment Simulator (JULES) are evaluated offline against multi-year flux observations in the densely built-up city centre of London and in suburban Swindon (UK): (i) the 1-tile slab model, used in climate simulations; (ii) the 2-tile canopy model MORUSES (Met Office–Reading Urban Surface Exchange Scheme), used for numerical weather prediction over the UK. Offline, both models perform better at the suburban site, where differences between the urban schemes are less pronounced due to larger vegetation fractions. At both sites, the outgoing short- and longwave radiation is more accurately represented than the turbulent heat fluxes. The seasonal variations of model skill are large in London, where the sensible heat flux in autumn and winter is strongly under-predicted if the large city centre magnitudes of anthropogenic heat emissions are not represented. The delayed timing of the sensible heat flux in the 1-tile model in London results in large negative bias in the morning. The partitioning of the urban surface into canyon and roof in MORUSES improves this as the roof tile is modelled with a very low thermal inertia, but phase and amplitude of the grid box-averaged flux critically depend on accurate knowledge of the plan-area fractions of streets and buildings. Not representing non-urban land cover (e.g. vegetation, inland water) in London results in severely under-predicted latent heat fluxes. Control runs demonstrate that the skill of both models can be greatly improved by providing accurate land cover and morphology information and using representative anthropogenic heat emissions, which is essential if the model output is intended to inform integrated urban services.

Keywords JULES · Land surface models · MORUSES · Surface-energy balance · Urban modelling

1 Introduction

As urbanisation levels and urban populations continue to grow (United Nations 2018), there is an increasing need for urban climate services (Baklanov et al. 2018). Climate models need to represent future extremes correctly, such as the intensity and frequency of heat waves, droughts or

floods. Urban areas can strongly exacerbate such events at the local scale due to their impact on the surface energy balance (SEB) and surface hydrology. Global climate models (GCM) have become increasingly high resolution, with frontier runs allowing horizontal resolutions (N1280 HadGEM3–PRIMAVERA (~10 km) Vidale et al. *in prep.*) that are similar to coarse resolution numerical weather prediction (NWP) applications (e.g. ECMWF; Holm et al. 2016). This means that large metropolitan areas like London cover several grid cells (assuming accurate land-cover information) and should be represented accordingly. Climate models can provide information on the likelihood and nature of (long-term) future extreme events under various forcing conditions. However, to develop appropriate plans and responses based on integrated urban services (WMO 2018), scenario runs should ideally be made at

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00704-020-03294-1>) contains supplementary material, which is available to authorized users.

✉ Denise Hertwig
d.hertwig@reading.ac.uk

Extended author information available on the last page of the article.

NWP resolution. To capture the impact of cities on surface–atmosphere interactions, urban land-surface models (ULSM) are used within NWP and climate models. ULSMs provide a solution to the urban surface-energy balance (USEB; e.g. Oke et al. 2017):

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

where $Q^* = (K^\downarrow - K^\uparrow) + (L^\downarrow - L^\uparrow)$ is the net all-wave radiation composed of the incoming (\downarrow) and outgoing (\uparrow) shortwave (K) and longwave (L) radiation components. Q_F is the anthropogenic heat flux, and Q_H and Q_E the turbulent sensible and latent heat fluxes. ΔQ_S is the net storage heat flux that includes the storage of heat within the urban fabric (thermal inertia, CdT_*/dt , where C is the areal capacity and T_* the surface temperature; Porson et al. 2010a), and the ground-heat flux. Net horizontal advection of heat or moisture is included on the right side of Eq. 1 if the ULSM is coupled to an atmospheric model.

1.1 Representation of urban processes

Urban processes are represented in atmospheric models with varying levels of morphological and physical

complexity depending on computational resources and resolution (Fig. 1). ULSMs used in NWP or climate simulations parameterise the effects of urban environments on surface fluxes without representing buildings explicitly. Numerous ULSMs exist (Grimmond et al. 2009) using different approaches to model sub-grid scale processes (Grimmond et al. 2010; Grimmond et al. 2011). With the ‘tile’ approach, urban areas are treated separately and in parallel with other surfaces in the grid box (e.g. vegetation or water). Simpler modelling approaches modify vegetation schemes (e.g. Best 2005) by using bulk surface parameters that are representative of impervious, urbanised surfaces (e.g. increased heat capacity and surface roughness). More complex models use parameterisations of processes in simple street canyons, for which morphology parameters can vary spatially (e.g. Masson 2000; Martilli et al. 2002; Porson et al. 2010a). A key problem at this scale is the description of the subgrid-scale heterogeneity of the urban surface in the models (Barlow et al. 2017).

Increasing the level of detail with which urban processes are modelled becomes important as grid resolutions approach the urban grey-zone regime ($\mathcal{O}(10 - 100 \text{ m})$) both in terms of physics (turbulent exchange is now partially resolved) and buildings (morphology variations from one grid box to the next can be significant). Neighbourhood to building-scale simulations feature building-dominant

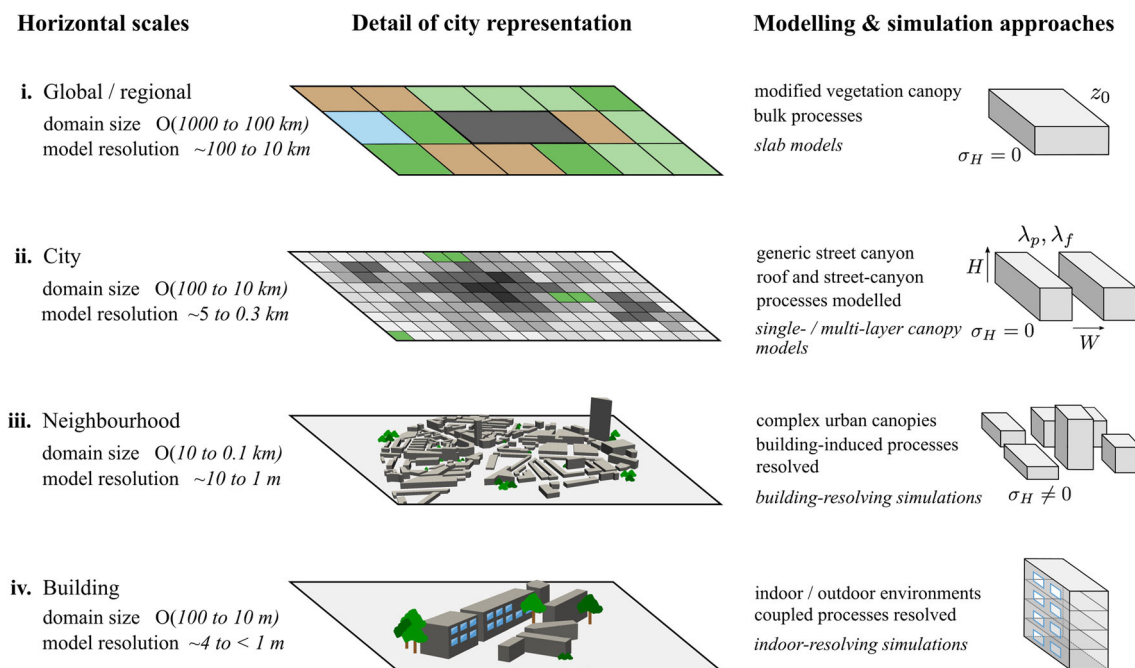


Fig. 1 Representation of urban areas at different spatial scales (domain sizes from $\mathcal{O}(1000 \text{ km})$ to $\mathcal{O}(10 \text{ m})$) and model resolutions together with prevailing modelling approaches, with H , mean building height;

σ_H , standard deviation of building heights; W , street-canyon width; z_0 , roughness length for momentum; λ_p , plan-area fraction; λ_f , frontal-area fraction

processes that need to be directly resolved and coupling of indoor–outdoor environments becomes possible.

1.2 Focus of this study

The choice of ULSM for a specific application is often constrained by practical considerations (e.g. Best 2006a), for example the computational resources available. With increasing model complexity, the demands on spatially varying ancillaries increase. Land-cover and building morphology/material information or population density and energy-use data to represent anthropogenic heat emissions may not be readily available at the spatial/temporal scales of interest.

The Met Office Unified Model (UM) is routinely run on regional to global scales in both climate mode (HadGEM) and NWP setups using different complexity ULSMs available in the Joint UK Land Environment Simulator (JULES; Best et al. 2011; Clark et al. 2011). The UM–JULES modelling suite has many international users beyond the UK, e.g. meteorological services in Australia, Singapore and South Korea, and an increasing demand for its application over highly urbanised areas of the world such as China and India. While HadGEM simulations currently use the simple one-tile (1T) urban scheme in JULES (Best 2005), since March 2016 the regional UKV NWP model (1.5-km horizontal resolution over the UK) is run operationally with the urban-canopy model MORUSES (Met Office–Reading Urban Surface Exchange Scheme; Porson et al. 2010a, 2010b; Bohnengstengel and Hendry 2016).

In this study, we focus on the following questions:

1. What are the implications of ULSM options, together with resolution and accuracy of ancillary information, in current UM–JULES modelling environments used for NWP and climate simulations?
2. How can current urban modelling capabilities in UM–JULES be improved, especially with regard to using the model output for integrated urban services?

To address these questions, the JULES ULSMs are evaluated offline against multi-year flux observations in the UK in two different urban settings: (i) the high-density, high-rise city centre of London (King’s College London Strand campus; KCL); and (ii) the suburban town of Swindon, located ~115 km west of London. The offline runs use the JULES configurations of (i) the operational UKV NWP setup and (ii) Global Land 7.0 (GL7.0; Walters et al. 2019; Wiltshire et al. 2020) used in HadGEM3 climate simulations. This is the first time that the JULES ULSMs are evaluated together over multi-year periods, allowing the

study of seasonal variations in model performance under a range of different conditions.

2 Urban models in JULES

JULES is a community land-surface model that uses a tile approach to represent the subgrid-scale heterogeneity of land cover (Best et al. 2011). This allows no interaction between tiles below the first atmospheric model level, where the total (grid box-averaged) surface fluxes are obtained by weighting the tile variables by their plan-area fractions (Best et al. 2006).

Urban areas are represented in JULES either using the (Best 2005) single urban tile scheme (Best 1T) or using a 2-tile (2T) approach that splits the impervious urban surface into a roof and a street canyon. For the 2T approach, JULES has two options: (i) a simple 2T slab scheme (Best 2T; Best et al. 2006) and (ii) MORUSES (Porson et al. 2010a; 2010b).

2.1 Representation of the USEB

The Best 1T and Best 2T approaches solve the SEB of a modified vegetation canopy that interacts with the first soil level via longwave radiative coupling. Urban bulk parameters for the canopy (Best 1T) or the roof/canyon combination (Best 2T) are prescribed (urban albedo α_U , urban emissivity ε_U , bulk areal heat capacity C_U , roughness length for momentum z_0 and the ratio z_h/z_0 , where z_h is the roughness length for heat). The maximum available surface water mass and infiltration rates are reduced to limit the evaporation over the impervious tile (Best 2005). In standard UM–JULES configurations, urban parameters for the Best models are the same for each city and across a city. While this is not an intrinsic model limitation, allowing for spatial variations of model parameters would require significant code changes in JULES.

MORUSES includes several modifications of the above: (i) bulk parameters for the street-canyon tile (canyon albedo α_c , emissivity ε_c and areal heat capacity C_c) are calculated from canyon geometry and prescribed material properties for roads and walls (material albedo α , emissivity ε , volumetric heat capacity c , thermal diffusivity k); (ii) z_0 and z_h are calculated explicitly (Section 3.3); and (iii) spatially varying urban morphology data can be provided (Bohnenstengel et al. 2011). The presence of a street canyon (building height H , street width W) in MORUSES allows modelling of short- and longwave radiative trapping. Turbulent scalar fluxes are computed through a resistance network that responds to canyon

geometry, modified from Harman et al. (2004). Three flow regimes are used: (i) isolated-building regime ($H/W \leq 1/3$), (ii) wake interference ($1/3 < H/W \leq 2/3$) and (iii) skimming flow ($H/W > 2/3$). The net storage heat flux (ΔQ_S) depends on thermal material properties and canyon geometry (Porson et al. 2010a). The canyon/roof tiles are conductively/radiatively coupled to the ground. For the canyon tile, only the road facet is coupled to the soil (not the walls). The insulation behaviour of the roof tile can be controlled via a prescribed thermal diffusivity or by using a very small roof thickness, Δz_f (Porson et al. 2010b).

2.2 Treatment of urban vegetation

UM–JULES typically uses five plant functional types (PFT) in separate tiles: broadleaf tree, needleleaf tree, C3 grass, C4 grass, shrubs. Urban vegetation is modelled using these vegetation tiles and not as part of the urban SEB. As a result, processes in built and vegetated areas do not interact below the first model level and Q_E only originates from urban tiles after rainfall. Shortcomings of this approach have been discussed by Grimmond et al. (2011) and Loridan and Grimmond (2012a, 2012b).

Prescribing urban anthropogenic moisture sources (e.g. irrigation) in JULES is not possible currently (Best and Grimmond 2016). While this is not critical for the evaluation sites considered in this study, effects of urban irrigation can be important in some cities (e.g. Ao et al. 2018; Dou et al. 2019).

2.3 Anthropogenic heat emissions

Currently, JULES (vn5.4) has two options for the emission of Q_F over the urban tile(s): (i) effectively none (0 W m^{-2}) and (ii) fixed values (JULES– Q_F) derived from the Digest of United Kingdom Energy Statistics (DUKES 2003). The latter is used operationally in the UKV and represents a UK-averaged monthly mean Q_F for 1995–2003 adjusted to the fraction dissipated over impervious/built areas. Consistent with this, when Q_F is added to the SEB in JULES it is only applied to the urban tile(s) and set to zero elsewhere. Hence, there is spatial variability at the grid scale as it is proportional to the urban fraction in the grid box. However, there are no local adjustments that account for differences between city centres and suburban areas, which arise from changes in building type and volume, population density and traffic volume (e.g. Sailor 2011; Iamarino et al. 2012). The minimum (unscaled; i.e. 100% urban) monthly JULES Q_F for the UK occurs in August ($\sim 17 \text{ W m}^{-2}$), the maximum in December ($\sim 26 \text{ W m}^{-2}$; Fig. 2).

Figure 2 shows JULES– Q_F together with results from the GreaterQf model (GQF; Iamarino et al. 2012; Gabey

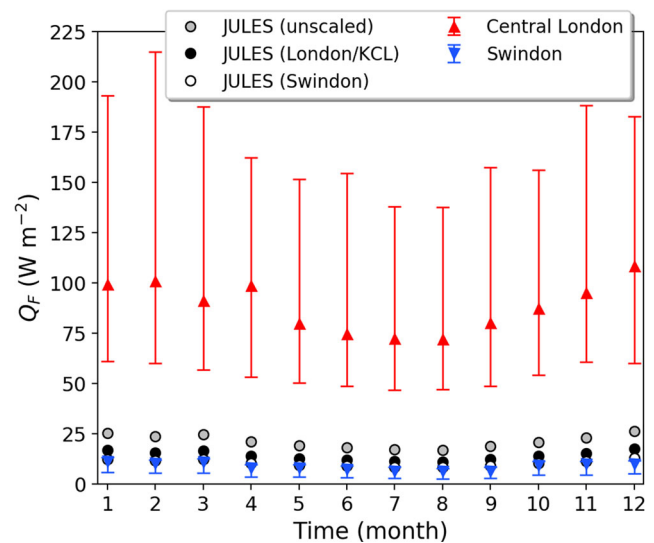


Fig. 2 Monthly anthropogenic heat flux (Q_F) specified in JULES (both unscaled and adjusted to urban fractions in London/Swindon), together with GQF output for central London (Kotthaus and Grimmond 2014a) and inventory heat emissions for suburban Swindon (Ward et al. 2013) aggregated as monthly medians plus inter-quartile range

et al. 2019) for central London (January 2011–December 2013; Kotthaus and Grimmond 2014a), and heat emissions for Swindon derived from inventory data (January 2011–September 2013; Ward et al. 2013). To compare with JULES, the hourly Q_F data for London and Swindon are aggregated into monthly medians. JULES– Q_F is shown both in its unscaled (100% urban) version and scaled to the urban tile proportions at the two sites (66% for London; 49% for Swindon; Table 1). The scaled values are representative of those used in the simulation.

In MORUSES, Q_F is either released only in the canyon or is evenly distributed among canyon and roof. The latter is the UKV default. The prescribed monthly values are linearly interpolated to the model time step, but diurnally constant. Hence, variations of Q_F related to sleep/work patterns, different energy demands on weekdays/weekends, public holidays or as a result of daylight-saving time, day-to-day or temperature conditions (Grimmond 1992; Sailor and Vasireddy 2006; Gabey et al. 2019) are not reflected. Using a fixed daily value can result in too large/low Q_F releases during the night/day (Grimmond et al. 2010).

3 Offline JULES–USLM configurations

JULES is run offline over 3 years (January 2011–December 2013) at an hourly time step using surface observations in central London and suburban Swindon as forcing data (Section 4; Table 5a). A modified version of JULES vn5.2 is

Table 1 Plan-area land-cover fractions used in the offline runs for (a) London/KCL and (b) Swindon

	BTr	NTr	C3Gr	C4Gr	Shr	Lake	Soil	Can	Roof	Urb
(a) London/KCL										
CTRL										
— Region T	0.10	0.0	0.03	0.0	0.0	0.21	0.0	0.29	0.37	0.66
— Region A	0.09	0.0	0.04	0.0	0.0	0.0	0.0	0.36	0.51	0.87
— Region B	0.12	0.0	0.0	0.0	0.0	0.38	0.0	0.20	0.30	0.50
— Region C	0.06	0.0	0.0	0.0	0.0	0.59	0.0	0.19	0.16	0.35
— Region D	0.09	0.0	0.0	0.0	0.0	0.31	0.0	0.36	0.24	0.60
UKV (west)	0.0	0.0	0.014	0.0	0.013	0.014	0.014	0.54	0.41	0.95
UKV (east)	0.0	0.0	0.02	0.0	0.02	0.14	0.03	0.58	0.21	0.79
HAD	0.0	0.0	0.0	0.0	0.0	0.02	0.0	—	—	0.98
(b) Swindon										
CTRL	0.08	0.01	0.36	0.0	0.0	0.0	0.06	0.33	0.16	0.49
UKV	0.01	0.0	0.49	0.0	0.12	0.0	0.07	0.26	0.05	0.31
HAD	0.01	0.01	0.62	0.05	0.02	0.0	0.06	—	—	0.23

In London, fractions for CTRL are given for the four regions A–D (Fig. 4) and the complete 500-m radius area (region T). For non-urban land-cover fractions standard vegetation tiles are used ('BTr': broadleaf tree; 'NTr': needleleaf tree; 'C3/C4Gr': C3/C4 grass, 'Shr': shrubs) together with inland water ('Lake') and bare soil. While MORUSES requires fractions for canyon ('Can') and roof, only the total urban ('Urb') plan area ('Urb' = 'Roof' + 'Can') is needed in Best 1T. For sources of the data see text

used that includes a fix enabling the radiative coupling of the MORUSES roof tile to the soil. This fix is part of the JULES release from vn5.4 and is expected to be included in the next available operational UKV release cycle. We compare three configurations (cf. Appendix 1):

1. HadGEM: JULES–GL7.0 science configuration (Walters et al. 2019; Wiltshire et al. 2020) and ~10 km resolution ancillaries of the N1280 HadGEM3–PRIMAVERA GCM system; using Best 1T (hereafter labelled: HAD–B)
2. UKV: Operational JULES–UKV science configuration (OS40) and 1.5-km resolution UKV ancillaries (OS41); using MORUSES (UKV–M)
3. Control: Operational JULES–UKV (OS40) together with high-resolution (1 m) land cover, building morphology and roughness data and the JULES leaf-phenology model (Cox 2001; Clark et al. 2011); using MORUSES (CTRL–M) and Best 1T (CTRL–B)

It is noted that while the CTRL–B runs use JULES–UKV, the SEB statistics remain mostly unchanged if JULES–GL7.0 is used with urban parameters and Q_F options kept the same (cf. seasonal diurnal statistics in the Online Resource Figs. S4, S5). Small differences stem from the hydrology and inland-water parameters (Appendix 1). Hence, 'CTRL–B v. HAD–B' (as 'CTRL–M v. UKV–M') differences are not from the model science, but reflect the impact of ULSM parameter choices, treatment of Q_F and differences in prescribed land cover.

3.1 Land-cover information

The operational UKV land-cover ancillaries are based on ~25-m resolution 1990 UK land-cover data (CEH 1990). The HadGEM3–PRIMAVERA ancillaries are derived from global ~1-km resolution data of the International Geosphere–Biosphere Project (IGBP; reference years 1995/1996; Walters et al. 2019, and references therein).

Figure 3a shows an urban versus non-urban binary land-cover classification over the south-east UK from the Global Urban Footprint (GUF) project that used TerraSAR–X/TanDEM–X radar data (~12-m resolution, reference year 2011; Esch et al. 2011; Esch et al. 2017; Esch et al. 2018).

The spatial representativeness of tile fractions for UKV–M and HAD–B varies with grid resolution of the atmospheric model (1.5 km and 10 km, respectively). The fraction of urban land cover (f_{urban}) in the UKV (Fig. 3b) and HadGEM3 (Fig. 3c) varies between the Greater London area and the surrounding urban regions. The coarser resolution HadGEM3–PRIMAVERA ancillaries (Fig. 3c) may make the outskirts or small urban areas (like Swindon) appear more vegetated. Urban fractions for Greater London, on the other hand, are over-estimated compared with the CEH-based UKV ancillaries (Fig. 3b; cf. Bohnenstengel and Hendry 2016). Highly urban coastal land tiles (Fig. 3c) are likely artefacts from the land-sea mask overlapping with only a small fraction of land with large f_{urban} . Affected grid cells are mostly sea, so the effect on the grid box results is small.

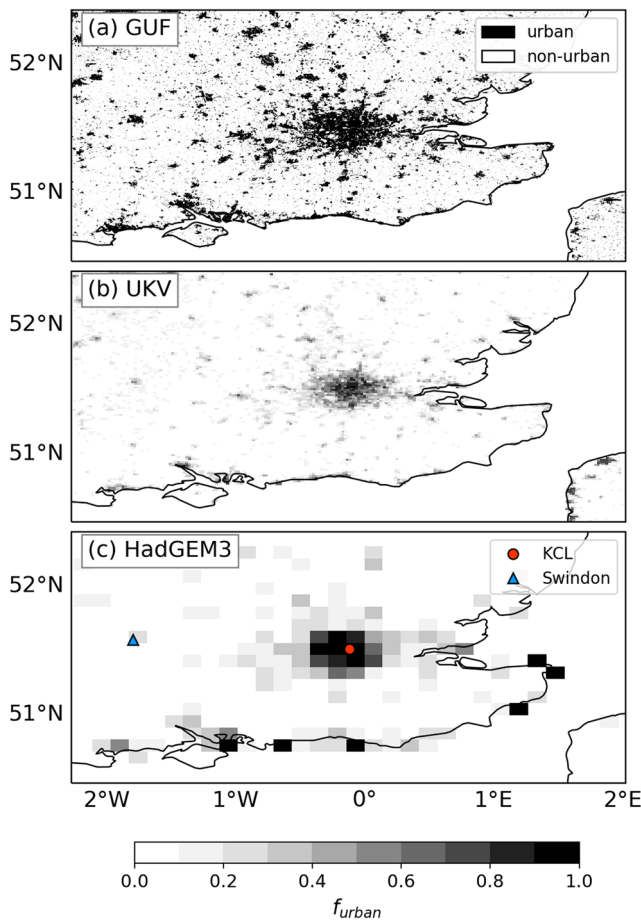


Fig. 3 Urban regions in the south-east UK as derived from (a) GUF (Esch et al. 2011; Esch et al. 2017; Esch et al. 2018) and urban land-cover fractions (f_{urban}) used in (b) the UKV and (c) HadGEM3-PRIMAVERA. See text for details on resolution and reference years

The land cover for the HAD-B and UKV-M runs is extracted from grid boxes containing the observation sites. In the HadGEM3 ancillaries, Swindon is represented by a single urbanised grid box ($f_{\text{urban}} = 0.23$, Table 1b). The KCL site in central London is located in a grid box with a very high urban fraction ($f_{\text{urban}} = 0.98$; Fig. 3c). In the UKV ancillaries, KCL sits at the border of two grid boxes (UKV west/east; Fig. 4) with quite different land cover (Table 1a). UKV-M at KCL is run for both grid boxes as the resolution accuracy of the computational mesh ($\mathcal{O}(10\text{ m})$) means that the site can be situated in either.

Land-cover and building-morphology parameters for the CTRL runs are for a 500-m radius area around each site, covering the source areas of the turbulent fluxes and radiation instruments (Ward et al. 2013; Kotthaus and Grimmond 2014b). For Swindon, the CTRL runs use the land cover reported by Ward et al. (2013) and Ward et al. (2016). For London/KCL, the land cover is derived from high-resolution (1m) GIS data using the Urban Multi-scale Environmental Predictor (UMEP; Lindberg et al. 2018). At

this site, the land cover varies strongly by wind direction. The River Thames dominates the area to the south, whereas the north is densely built-up (Fig. 4). For the CTRL runs, the total 500-m radius area (region T) is used, together with four additional segments (regions A, B, C and D; Fig. 4) that are sub-sampled based on wind direction for the evaluation of the turbulent heat fluxes (Section 5.1). At Swindon, there is slightly more vegetation to the north-east and slightly more impervious surfaces to the south-west (Fig. 2 in Ward et al. 2013), but land-cover variations are not as marked as in London.

3.2 Building morphology

The input parameters required for MORUSES are as follows: mean building height (H), canyon aspect ratio (H/W) and canyon-width ratio (W/R), where W is the street-canyon width and R is the combined length of canyon and roof. The ratios can be related to the plan and frontal-area fractions (λ_p, λ_f ; Porson et al. 2010a) via:

$$\frac{W}{R} = 1 - \lambda_p \quad (2)$$

$$\frac{H}{W} = \frac{\pi}{2} \frac{\lambda_f}{(1 - \lambda_p)}, \quad (3)$$

where λ_p is the proportion of the urban tile occupied by buildings (i.e. the roof fraction), so that W/R is the fraction of non-building impervious (street canyon) surface.

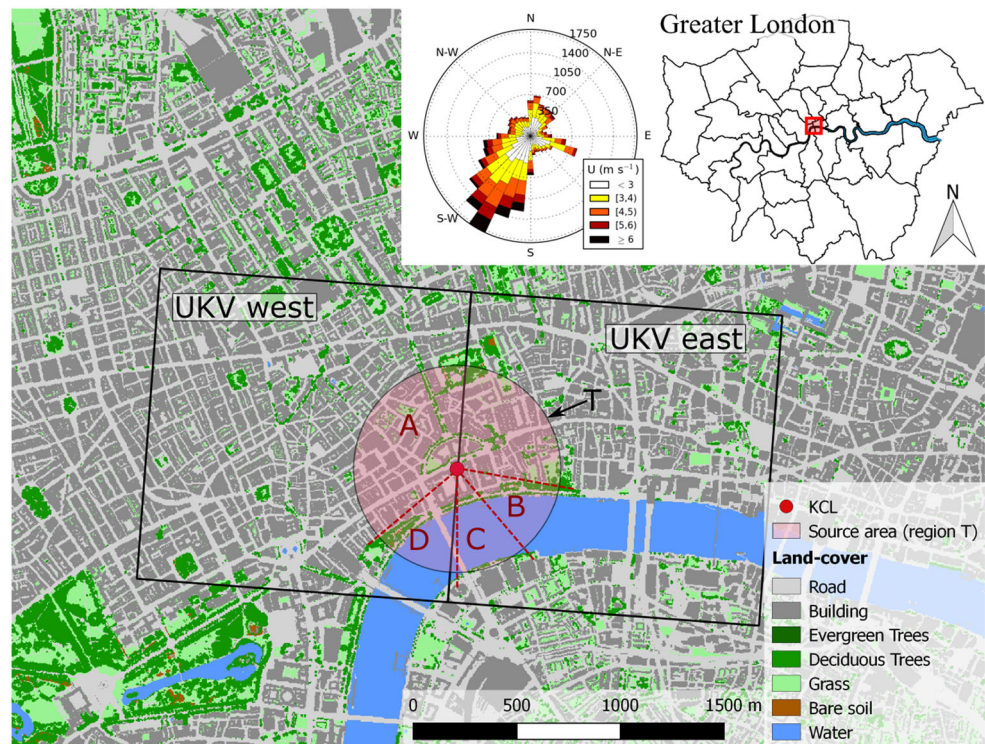
The operational UKV morphology ancillaries are generated from empirical polynomial relations derived from high-resolution ($\mathcal{O}(1\text{ m})$) 3D London morphology data (Virtual London dataset; Evans et al. 2006) that form the basis of the parametrisations of λ_p, λ_f (Bohnenstengel et al. 2011) and H (Gilham et al. 2019) with the urban fraction in the grid box, f_{urban} . Given the $\sim 1\text{-km}$ resolution for NWP, a 50-m radius smoothing filter removed small-scale land-cover variations (details in Bohnenstengel et al. 2011). As the empirical relations were only derived for one city, their use elsewhere should be assessed or the MORUSES morphology parameters should be derived from local data.

For the CTRL runs in London, H, λ_p and λ_f are derived directly from 1-m resolution GIS data. For consistency, Eq. 3 is used to obtain H/W despite UMEP being able to derive the aspect ratio directly (Lindberg et al. 2018). For the Swindon site, the morphology and roughness characteristics reported by Kent et al. (2018) are used.

3.3 Roughness parameters

Roughness and morphology parameters for the offline runs are summarised in Tables 2a, 3a and 4. HAD-B uses the

Fig. 4 Land-cover distribution in central London around the KCL site (red dot), with wind rose (January 2011–December 2013; inset) and the UKV west/east grid boxes. The 500-m radius area (large circle; region T) contains the 80% source area of turbulent heat fluxes for the site (neutral and unstable conditions). Labels (A–D) indicate the four sectors defined to sub-sample the CTRL runs. CTRL–T is for the complete circle



default Best-1T z_0 of 1 m. In all runs with the Best-1T model, z_h/z_0 uses the fixed default value of 10^{-7} (Best et al. 2006).

In MORUSES, z_h is calculated from the bulk resistances to scalar transport of the roof and canyon tiles (Porson et al. 2010a, and Eq. 10 in Appendix 2). The canyon resistances vary in response to H/W (Section 2.1). In the isolated-building flow regime ($H/W \leq 1/3$), the upper limit of z_h in MORUSES is fixed to $0.1z_{0m}$, where $z_{0m} = 0.05$ m is the prescribed material roughness length for momentum.

The resistance network is driven by the flow at the roof and canyon top, obtained by extrapolating from the model-forcing level without correcting for stability effects and assuming that the logarithmic layer extends down to the building top (Harman et al. 2004). The roughness parameters, z_d and z_0 , needed for this calculation can be obtained from the Macdonald et al. (1998) morphometric method available in MORUSES (Eqs. 8 and 9; Appendix 2) or can be prescribed. For the CTRL runs, the Macdonald z_0 and z_d are calculated externally and then prescribed. Value differences between UKV–M and CTRL (Table 2a) are caused by the differences in resolution and method of derivation of the underlying morphology data.

Modifications to the Macdonald method to account for building-height variability, like the Kanda et al. (2013) approach, have been shown to provide better predictions of wind speeds above cities (Kent et al. 2017; Kent et al. 2018a). While in the Macdonald approach z_d scales with H , Kanda's modification accounts for the maximum building

height within z_d . For the 500-m radius area around the London site (region T; Fig. 4), the Kanda $z_0 = 2.3$ m and $z_d = 30$ m (i.e. approximately 10 m above H) are 0.5 m and 17 m greater, respectively, than the Macdonald values (Table 2a).

Being able to reflect the impact of increased drag on the flow is crucial if the ULSM is applied in environments with tall buildings and/or a large heterogeneity of building heights. Although the Macdonald model available in MORUSES effectively limits the displacement height such that $z_d \in [0, H]$ (Eq. 8), the facility exists to supply other data as ancillaries. Conceptually, JULES estimates the surface exchange at a height of $z_d + z_0$, representing an effective surface in the model. This means that the effective height of the street canyon in MORUSES is $H - z_d - z_0$. If $H \leq z_d + z_0$, as can be the case in high-rise city centres, the velocity at the canyon top currently defaults to zero. MORUSES requires further development in order to make use of parameterisations that enter this regime. For the roof tile, the criterion is slightly less strict with $1.1H \leq z_d + z_0$.

3.4 Radiation and thermal controls

Urban radiative and thermal parameters used in this study are summarised in Table 2b and c. UKV–M prescribes the same radiative and thermal urban fabric properties across UK cities, assuming asphalt roads, brick walls and clay roofs (Porson et al. 2010a; Bohnenstengel et al. 2011). In London, the CTRL–M radiative parameters are adjusted to

Table 2 Overview of parameters used for JULES (Best JT and MORUSES) in central London (KCL; 'L') and Swindon ('S')

Parameter	Units	CTRL-M		CTRL-B		UKV-M		HAD-B	
		L	S	L	S	L _w	L _E	L	S
(a) Morphology and roughness									
Mean building height	<i>H</i>	19.81	4.7	—	—	13.03	7.99	—	5.44
Canyon fraction	<i>W/R</i>	0.44	0.67	—	—	0.57	0.73	—	0.85
Height-to-width ratio	<i>H/W</i>	1.18	0.28	—	—	0.95	0.40	—	0.18
Roughness length	<i>z₀</i>	1.76	0.35	1.76	0.35	0.79 ^a	0.70 ^a	1.0	0.50 ^a
Displacement height	<i>z_d</i>	13.19	2.0	—	—	9.11 ^a	4.09 ^a	—	1.74 ^a
Material roughness length	<i>z_{0m}</i>	0.05	0.05	—	—	0.05	0.05	—	0.05
Radiative properties									
Bulk urban albedo	α_U	0.09 ^{b,c}	0.11 ^{b,c}	0.11	0.11	0.12 ^{b,c}	0.11 ^{b,c}	0.18	0.10 ^{b,c}
Wall albedo	α_w	0.13	0.375	—	—	0.375	0.375	—	0.375
Road —	α_r	0.11	0.08	—	—	0.08	0.08	—	0.08
Roof —	α_f	0.13	0.18	—	—	0.18	0.18	—	0.18
Canyon —	α_c	0.04 ^{a,c}	0.08 ^{a,c}	—	—	0.08 ^{a,c}	0.08 ^{a,c}	—	0.08 ^{a,c}
Bulk urban emissivity	ε_U	0.94 ^b	0.96 ^b	0.97	0.97	0.97 ^b	0.96 ^b	0.97	0.96 ^b
Wall emissivity	ε_w	0.91	0.875	—	—	0.875	0.875	—	0.875
Road —	ε_r	0.95	0.95	—	—	0.95	0.95	—	0.95
Roof —	ε_f	0.91	0.97	—	—	0.97	0.97	—	0.97
Canyon —	ε_c	0.98 ^a	0.96 ^a	—	—	0.97 ^a	0.96 ^a	—	0.96 ^a
Thermal properties									
Bulk areal heat capacity	<i>C_U</i>	0.42 ^b	0.21 ^b	0.28	0.28	0.33 ^b	0.26 ^b	0.28	0.23 ^b
Canyon —	<i>C_c</i>	0.90 ^a	0.29 ^a	—	—	0.54 ^a	0.34 ^a	—	0.26 ^a
Roof —	<i>C_f</i>	0.04 ^a	0.04 ^a	—	—	0.04 ^a	0.04 ^a	—	0.04 ^a
Wall vol. heat capacity	<i>c_w</i>	2.11	1.37	—	—	1.37	1.37	—	1.37
Road —	<i>c_r</i>	1.94	1.94	—	—	1.94	1.94	—	1.94
Roof —	<i>c_f</i>	1.77	1.77	—	—	1.77	1.77	—	1.77
Wall thermal diffusivity	<i>k_w</i>	0.72	0.61	—	—	0.61	0.61	—	0.61
Road —	<i>k_r</i>	0.38	0.38	—	—	0.38	0.38	—	0.38
Roof —	<i>k_f</i>	0.47	0.47	—	—	0.47	0.47	—	0.47
Wall thickness	Δz_w	0.14 ^a	0.13 ^a	—	—	0.13 ^a	0.13 ^a	—	0.13 ^a
Road —	Δz_r	0.10 ^a	0.10 ^a	—	—	0.10 ^a	0.10 ^a	—	0.10 ^a
Roof —	Δz_f	0.02 ^d	0.02 ^d	—	—	0.02 ^d	0.02 ^d	—	0.02 ^d

^aComputed internally by MORUSES at run time^bComputed externally as the weighted average of canyon and roof parameters^cVaries with solar zenith angle. Value reported is the summer (JJA) median at solar noon^dEffectively prescribed by using the 'thin roof' option; otherwise computed at run timeCTRL parameters at KCL are for a 500-m radius area (region T; Table 1, Fig. 4); parameters for sectors A–D are given in Table 3. UKV-M parameters at KCL are for west and east grids ('L_w', 'L_E'). For sources of the data, see text

Table 3 As Table 2, but for London/KCL CTRL–M runs for sectors A–D (Fig. 4)

Parameter		Units	CTRL–M			
(a)	Morphology and roughness		A	B	C	D
	Mean building height	H m	21.3	22.5	21.6	22.8
	Canyon fraction	W/R —	0.41	0.40	0.54	0.60
	Height-to-width ratio	H/W —	1.66	0.66	0.36	0.58
	Roughness length	z_0 m	1.1	3.1	3.3	3.4
	Displacement height	z_d m	15.9	11.1	7.0	8.7
	Material roughness length	z_{0m} m	0.05	0.05	0.05	0.05
(b)	Radiative properties		A	B	C	D
	Bulk urban albedo	α_U —	0.09 ^{a,b}	0.10 ^{a,b}	0.10 ^{a,b}	0.09 ^{a,b}
	Canyon albedo	α_c —	0.03 ^{c,b}	0.06 ^{c,b}	0.08 ^{c,b}	0.06 ^{c,b}
	Bulk urban emissivity	ε_U —	0.94 ^a	0.93 ^a	0.94 ^a	0.95 ^a
	Canyon emissivity	ε_c —	0.98 ^c	0.97 ^c	0.96 ^c	0.97 ^c
(c)	Thermal properties		A	B	C	D
	Bulk areal heat capacity	C_U MJ K ⁻¹ m ⁻²	0.51 ^a	0.26 ^a	0.26 ^a	0.34 ^a
	Canyon —	C_c MJ K ⁻¹ m ⁻²	1.18 ^c	0.59 ^c	0.41 ^c	0.54 ^c
	Roof —	C_f MJ K ⁻¹ m ⁻²	0.04 ^c	0.04 ^c	0.04 ^c	0.04 ^c

^aComputed externally as the weighted average of canyon and roof parameters

^bVaries with solar zenith angle. Value reported is the summer (JJA) median at solar noon

^cComputed internally by MORUSES at run time

Thermal and radiative material parameters are the same as for region T (cf. Table 2b, c). For sources of the data, see text

provide the best match to the observed bulk parameters reported by Kotthaus and Grimmond (2014a). For CTRL–M at Swindon, the default UKV parameters are representative of typical building materials at this site (Ward et al. 2013).

In MORUSES, the ‘thin-roof’ setting is used (UKV default), by which insulation is modelled through a very

small roof thickness (the minimum of $\Delta z_f = 0.02$ m and the calculated damping depth). Earlier comparisons with the ‘thick-roof’ case, for which $\Delta z_f \propto k_f^{0.5}$ (Eq. 45 in Porson et al. 2010a), showed that the ‘thin-roof’ option improves the phase of Q_H (Porson et al. 2010b).

HAD–B uses the default bulk values of $\alpha_U = 0.18$, $\varepsilon_U = 0.97$ and $C_U = 0.28$ MJ K⁻¹ m⁻². For CTRL–B, the bulk α_U and ε_U are derived from observations (Kotthaus and Grimmond 2014a; Ward et al. 2013), while C_U was left unchanged from the default HAD–B value.

In all runs, the (snow-free) albedos for bare soil ($\alpha = 0.11$) and inland water (JULES–UKV: $\alpha = 0.06$; JULES–GL7.0: $\alpha = 0.12$) are prescribed. For vegetated surfaces, JULES calculates the albedo as a function of leaf-area index (LAI; m² m⁻²; Best et al. 2011).

3.5 Vegetation controls

LAI and canopy height (H_c ; m) for UKV–M and HAD–B are based on the UKV and HadGEM3–PRIMAVERA ancillaries that vary spatially and by month. In this data source, LAI is effectively zero for all PFTs in highly urbanised grid boxes such as central London, although vegetation is present (Fig. 4). For UKV–M at Swindon, this also affects the tree and shrub tiles. The LAI ancillaries are based on the relatively coarse IGBP land-cover mapping (Fig. 3c), which for the UKV can lead to inconsistencies

Table 4 Ratio of roughness lengths for heat and momentum for the MORUSES runs at (a) London/KCL and (b) Swindon

		$\overline{z_h}/z_0$
(a) London/KCL	CTRL–M	
	— Region T	$6.5 \cdot 10^{-4}$
	— Region A	$7.2 \cdot 10^{-3}$
	— Region B	$2.8 \cdot 10^{-5}$
	— Region C	$2.4 \cdot 10^{-6}$
	— Region D	$3.8 \cdot 10^{-5}$
	UKV–M (west)	$7.4 \cdot 10^{-3}$
	UKV–M (east)	$3.5 \cdot 10^{-3}$
(b) Swindon	CTRL–M	$1.3 \cdot 10^{-2}$
	UKV–M	$2.8 \cdot 10^{-3}$

$\overline{z_h}$ is externally computed as the weighted average of the canyon and roof tile z_h . In CTRL–B and HAD–B the ratio is fixed to 10^{-7}

with the CEH land cover. The canopy height, however, is non-zero so that roughness effects of the vegetation are modelled.

In case of near-zero LAI, JULES treats the vegetated tiles effectively as bare-soil surfaces with zero evapotranspiration and minimised water-holding and interception capacity (Best et al. 2011). In this study, the affected vegetation fractions are low (0.01–0.12; Table 1), but even small amounts of vegetation can have a strong impact on the surface-energy partitioning in urban areas (Grimmond et al. 2010; Loridan and Grimmond 2012a; 2012b).

Another feature shared by the UKV and HadGEM3–PRIMAVERA ancillaries over the southern UK is an anomalous rise in LAI in December/January following a minimum in November. This affects all PFTs and stems from the underlying MODIS Terra 4-km LAI dataset (collection 5; 2005–2009). The newer MODIS (collection 6) products do not show this artefact over the same analysis period.

The CTRL runs use the JULES leaf-phenology model (Cox 2001; Clark et al. 2011) to simulate leaf-on/leaf-off periods in response to temperature changes and the UKV H_c ancillaries are prescribed.

3.6 Anthropogenic heat flux

The CTRL runs use the monthly median GQF results for London/KCL and inventory emissions for Swindon (Fig. 2), both with and without scaling to the urban fraction (cf. Section 2.3). The former option is consistent with the way in which the value is used within JULES and the latter option results in an underestimation of the heat supplied to the system with respect to the GQF and inventory emission values. This is because JULES assumes that Q_F is released only on the urban tile(s); thus, the grid value is correspondingly reduced. The GQF and inventory data, however, are for the complete area. Figures S1 and S2 in the supplementary Online Resource show results from using the scaled values of Q_F , consistent with the assumptions within JULES. CTRL–B in London (Fig. S1d) has a reduced Q_H MBE over all seasons and CTRL–M/B in DJF both benefit from the increased magnitude of Q_F when compared with the results using the unscaled values (Section 6.1.2; Fig. 5d), but the scaling leads to larger errors overall. However, there is a large uncertainty in the estimation of Q_F and it is unclear what a representative magnitude of Q_F ideally should be for the sites. Therefore, we discuss the CTRL runs with the unscaled Q_F in Section 6.

In UKV–M, the default JULES– Q_F values are used (Fig. 2), and Q_F is neglected in HAD–B, reflecting the JULES settings for the UKV and GL7.0.

3.7 Model spin-up and output

For each of the four JULES soil layers (0.1, 0.25, 0.65 and 2.0 m depths), soil temperature and soil wetness (mass of soil water as a fraction of water content at saturation) are initially prescribed. The run is considered ‘spun-up’ if the change in soil-moisture content (kg m^{-2}) is below 1% and the change in soil temperature below 0.2 K over a spin-up cycle. The first year of forcing data is used for the spin-up.

The soil states reached after the successful spin-up completion differ in response to the JULES configuration and land-cover ancillaries (cf. Appendix 3). The differences in soil states are maintained (qualitatively) throughout the evaluation period, indicating another difference between experimental setups. However, repeating the CTRL–M/B simulations with swapped initial conditions in a sensitivity test, in London results in negligible differences between the turbulent heat-flux statistics after the first 4 months or so of the simulation. Whereas in the more vegetated Swindon, differences persist over the first year with the largest impact in summer.

Results from the offline runs are a sample at the model time step (1 h). Seasonal and diurnal statistics are computed in a post-processing step, in which the model output is analysed only when observations are available (Section 4).

4 Observations

For model forcing and evaluation hourly averages measured in central London (KCL; 51.511°N, 0.117°W) and Swindon (51.585°N, 1.798°W) are used. The London/KCL data comprise measurements from two sites on the same roof (horizontal distance < 45 m; vertical offset 1.4 m), which operated consecutively during the study period (Kotthaus and Grimmond 2014a, b; Bjoerkegren et al. 2015). Following Ward et al. (2016), we define the measurement height as the height of the wind-speed measurements above ground level (agl), which is 10.6 m in Swindon and 49.6 m at KCL. The latter is the average height of both KCL sites.

The forcing height in JULES is the height of the instruments above the local z_d (Table 2a; KCL z_d for region T). For KCL, this is 36.4 m; for Swindon 8.6 m. At both sites, the JULES forcing heights for temperature/humidity ($z_{T,q}$) and wind ($z_{u,v}$) are equal.

Table 5 (Appendix 4) gives an overview of observed variables and instrumentation. Observed forcing variables (Table 5a) are K^\downarrow , L^\downarrow , air temperature (T_{air}), rainfall rate (RR), wind speed (U_h), specific humidity (q) and atmospheric pressure (P). Periods of snowfall and settled snow are insignificant at both sites during the evaluation

periods (Ward et al. 2016). The forcing data are gap-filled observations used in the offline evaluation of the SUEWS land-surface model for 01 January 2011–31 December 2013 (Ward et al. 2016).

The evaluation focuses on the outgoing radiation components (K^\uparrow , L^\uparrow) and the net all-wave radiation (Q^*), measured with radiometers, and the turbulent heat fluxes (Q_H , Q_E) derived from eddy-covariance measurements (Table 5b). Evaluation data are available at Swindon over a 2-year period (09 May 2011–25 April 2013) and in London for the entire 3-year simulation span. Occasional data gaps within these periods generally are of short duration. Detailed descriptions of site characteristics, instrumentation, data processing techniques and analyses are provided by Kotthaus and Grimmond (2014a, b) and Bjorkegren et al. (2015) for London/KCL and Ward et al. (2013) for Swindon.

5 Evaluation methods

5.1 Source-area sub-sampling

The JULES tiling scheme does not account for the distribution of land-cover features within a grid box. In order to reflect the difference in the eddy-covariance source-area characteristics around London/KCL (Fig. 4), the CTRL-M and CTRL-B results are created by sub-sampling the grid box Q_H and Q_E of the individual CTRL runs for sectors A to D based on the observed horizontal wind direction (θ). For the 3-year study period, 52.6% of samples fall into region A, 9.6% in B, 5.8% in C and 32% in D (cf. wind rose, Fig. 4).

The four source-area segments have large differences in land-cover characteristics (Table 1): region A ($0^\circ \leq \theta < 100^\circ$ and $230^\circ \leq \theta < 360^\circ$) is dominated by impervious surfaces with little vegetation; region C ($140^\circ \leq \theta < 180^\circ$) includes large fractions of the River Thames and regions B ($100^\circ \leq \theta < 140^\circ$) and D ($180^\circ \leq \theta < 230^\circ$) include water, impervious surfaces and urban park vegetation.

The source area of the radiometers is independent of θ , but shadowing fractions vary by time of day and seasonally. In London, the runs for region T are compared with the radiation observations.

5.2 Evaluation metrics

To quantify the model performance, predicted (P_i) and observed (O_i) variables are compared in terms of evaluation metrics. These include the mean absolute error:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N (|P_i - O_i|), \quad (4)$$

the mean bias error:

$$\text{MBE} = \frac{1}{N} \sum_{i=1}^N (P_i - O_i), \quad (5)$$

the normalised centred root-mean-square error:

$$\text{nRMSE} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N [(P_i - \bar{P}) - (O_i - \bar{O})]^2}}{\sigma_O} \quad (6)$$

and the correlation coefficient (Pearson's R):

$$R = \frac{\frac{1}{N} \sum_{i=1}^N [(P_i - \bar{P})(O_i - \bar{O})]}{\sigma_P \sigma_O}, \quad (7)$$

where N is the sample size and σ_O and σ_P are the standard deviations of observations and predictions, respectively.

Online evaluations of the UKV run with MORUSES and Best 1T noted different model behaviours based on the timing of turbulent heat fluxes in the morning (Warren et al. 2018). For the offline tests, the metrics are therefore computed for 24-h periods and for sections of the diurnal cycle. Local sunrise (SR), solar noon (SN) and sunset (SS) times are used to aggregate the samples at time t into one of four periods:

1. Morning (M): $\text{SR}+1 \text{ h} < t \leq \text{SN}$
2. Afternoon (A): $\text{SN} < t < \text{SS}-1 \text{ h}$
3. Transition (T): $\text{SR}-1 \text{ h} \leq t \leq \text{SR}+1 \text{ h}$ and $\text{SS}-1 \text{ h} \leq t \leq \text{SS}+1 \text{ h}$
4. Night (N): $\text{SS}+1 \text{ h} < t < \text{SR}-1 \text{ h}$

6 Offline ULSM evaluation results

6.1 Seasonal-diurnal USEB variability

SEB components are analysed in terms of median diurnal cycles of grid box averages (i.e. averaged over all tiles) in central London (Fig. 5; 3-year period) and suburban Swindon (Fig. 6; 2-year period) in winter (DJF), spring (MAM), summer (JJA) and autumn (SON). Differences in the JULES-ULSMs performance are overall more pronounced for London, as in Swindon the urban plan-area fraction is a much smaller part of the grid box and therefore the contribution to the fluxes is lower (Table 1). Hence, the Swindon results are to a much larger degree determined by processes in the vegetated canopies.

6.1.1 Radiation and storage-heat flux

The energy available for partitioning into the turbulent heat fluxes is determined from $Q^* + Q_F - \Delta Q_S$ (Eq. 1).

As Q_F is prescribed, first it is important to evaluate the outgoing radiation (K^\uparrow , L^\uparrow) and storage-heat flux. While the radiometer at London/KCL sees almost exclusively roofs, urban canyons and some trees, in Swindon K^\uparrow and L^\uparrow are to a large degree affected by vegetation.

Differences in prescribed bulk and material albedos (Table 2b) impact K^\uparrow (Figs. 5a, 6a). At both sites, HAD-B consistently shows the largest deviation from the observations. This can partly be attributed to the comparatively large default urban albedo used in JULES-GL7.0 ($\alpha_U = 0.18$). The different land-cover make-up for HAD-B compared with the observed source

area also contributes to the differences in the grid box K^\uparrow , particularly in London through a larger contribution from the urban tile ($f_{\text{urban}} = 0.98$). However, even when prescribing land cover that is representative of the radiometer source area (as in CTRL), the HAD-B error remains larger compared with the other runs (cf. Online Resource Figs. S4a, S5a), highlighting the impact of α_U . CTRL-B (Figs. 5a, 6a) uses the observed α_U of 0.11. In London, this lower bulk urban albedo together with the addition of 21% inland water ($\alpha = 0.06$) in CTRL-B (region T, Table 1a) greatly improves the agreement with the observations over 24 h (low MAE and |MBE|).

Fig. 5 Modelled and observed median seasonal diurnal cycles with inter-quartile range (shading) at London/KCL for (a) K^\uparrow , (b) L^\uparrow , (c) Q^* , (d) Q_H and (e) Q_E in winter (DJF), spring (MAM), summer (JJA) and autumn (SON). All quantities represent grid box averages. The sample size (N) of the hourly data pairs and the seasonal 24-h MAE / MBE (W m^{-2}) shown for (1) CTRL-M, (2) CTRL-B, (3.a) UKV-M (east), (3.b) UKV-M (west) and (4) HAD-B, with the lowest MAE and |MBE| highlighted (bold). See Section 5.2 for metrics and Table 2 for run parameters

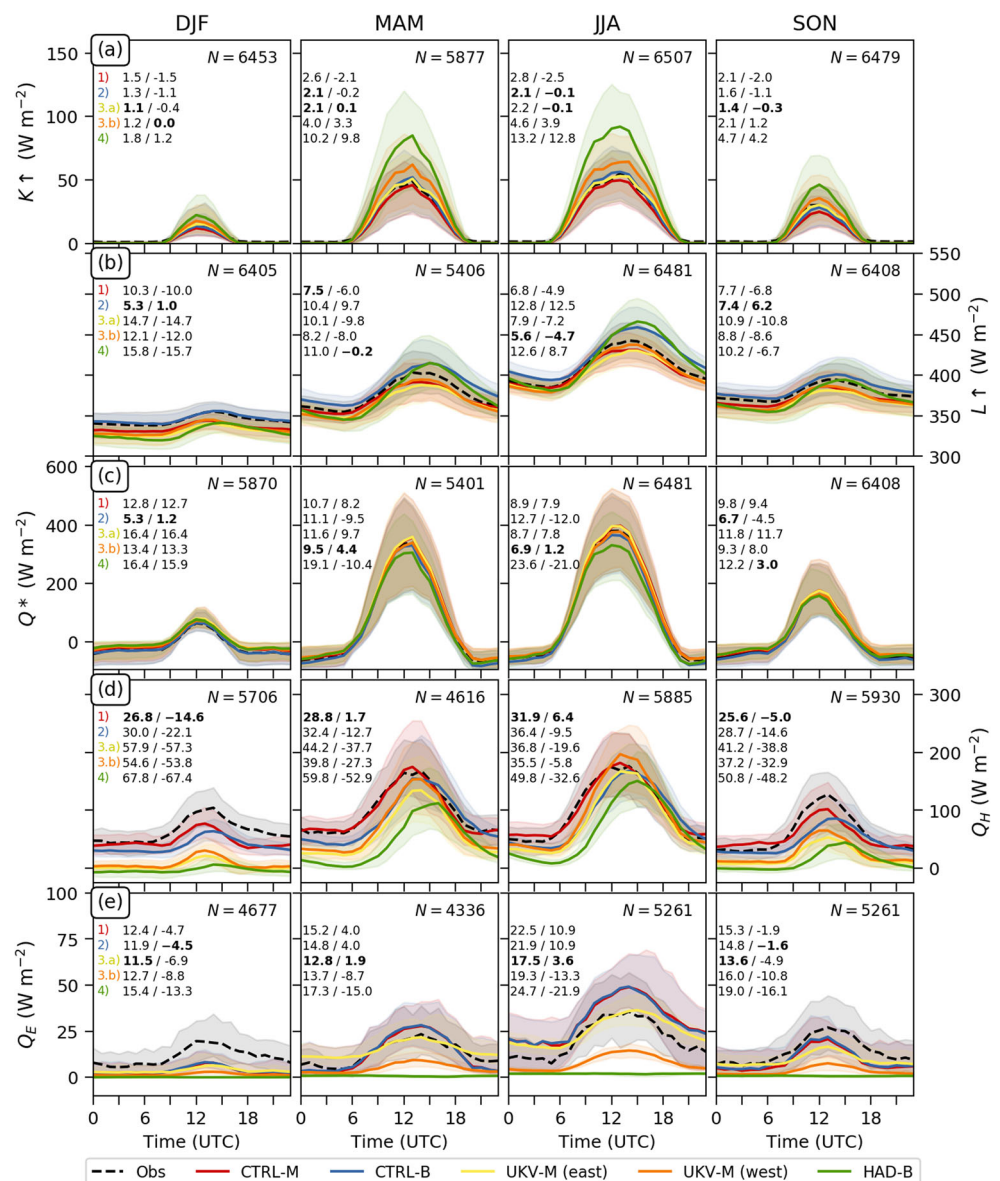
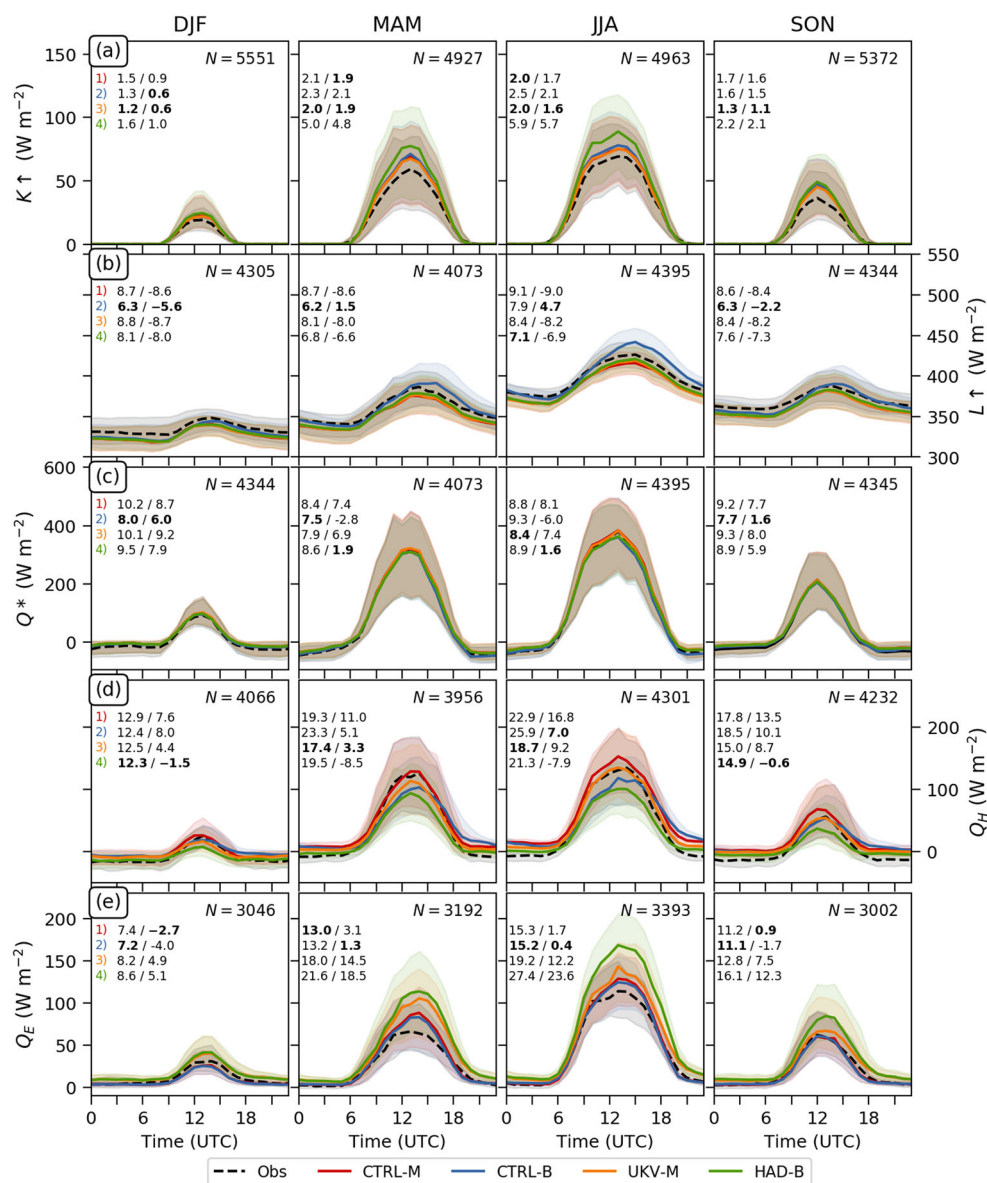


Fig. 6 As Fig. 5, but for Swindon with the seasonal 24-h MAE / MBE (W m^{-2}) for (1) CTRL-M, (2) CTRL-B, (3) UKV-M, (4) HAD-B



MORUSES assumes a (fixed) fraction of $K \downarrow$ is scattered by the sky and uses the Masson (2000) shadowing factors and multiple reflections within the canyon to obtain the canyon albedo (α_c). The latter depends both on H/W and on solar zenith angle (cf. Fig. 3 in Porson et al. 2010a), resulting in seasonal and diurnal variations. Depending on H/W and the difference of the wall and road albedos prescribed, α_c can reach a maximum or minimum at solar noon (cf. Figs. 6, 7 in Masson 2000). The former is the case if wall and road albedos are similar (e.g. CTRL-M in London; Online Resource Fig. S3a), while the latter regime arises when they are very different (default JULES-UKV values for α_r, α_w ; Online Resource Fig. S3b,c).

α_c reported in Table 2b is the JJA median at solar noon, when the albedo has the greatest impact as $K \downarrow$ is

generally at its peak. However, especially for the runs using the UKV-M material albedos, α_c shows strong seasonal-diurnal variability (Online Resource Fig. S3b,c). In London, CTRL-M (region T) computes a very low JJA canyon albedo of 0.04 (much lower than the roof albedo of 0.13) that varies only little by season or time of day. Using the default UKV material albedos, CTRL-M in Swindon and UKV-M at both sites have the same JJA α_c of 0.08. However, the differences between α_c for UKV-M east/west are larger in other seasons (cf. Online Resource Fig. S3b) as trapping effects in MORUSES are more pronounced when the sun is high, and hence are stronger in summer than in winter and stronger for larger H/W .

A bulk urban albedo, α_U , for the canyon-roof combination can be obtained as the weighted average of α_c (here the

JJA median) and α_f (Table 2b) to compare with the prescribed Best-1T value. While CTRL-M in London has an α_U of 0.09, the UKV-M runs result in bulk albedos of 0.11 (east grid) and 0.12 (west grid; Table 2b). For both UKV-M runs, the over-prediction of the grid box K^\uparrow is notably reduced compared with HAD-B (Fig. 5a). As in both HAD-B and UKV-M (west), the results are dominated by the urban tile(s) (98% and 95% total urban fractions, respectively); the bias reduction seen for the latter can be attributed to the lower α_U computed by MORUSES. Differences between the UKV-M east/west results can be explained by the contribution of the inland-water tile in the east-grid run (14%; Table 1a), which has a very low prescribed albedo (0.06 in JULES-UKV). In Swindon, the larger amplitude of K^\uparrow compared with London is better predicted by all models, with HAD-B again having the largest over-prediction attributable to the larger α_U (cf. Online Resource Fig. S5a).

In London, Best-1T (CTRL-B, HAD-B) has a notably larger grid box L^\uparrow (Fig. 5b) in spring and summer compared with the observations, with a ~ 1 -h delay of the peak and larger amplitudes throughout the afternoon and early night. In Swindon (Fig. 6b), this feature is present in the CTRL-B results ($f_{\text{urban}} = 0.49$), but not in HAD-B, for which the urban fraction is much lower (0.23) and the grid box L^\uparrow is dominated by vegetation. MORUSES captures better the phase of the observations, but shows a slight systematic under-prediction of L^\uparrow over 24 h at both sites and for all configurations.

In addition to the surface emissivity (Table 2b), L^\uparrow is controlled by the surface temperature (T_*), which is affected by the effectiveness of heat storage within the urban tile and hence by the areal heat capacity. For CTRL-B and HAD-B, the bulk urban C_U is fixed at $0.28 \text{ MJ K}^{-1} \text{ m}^{-2}$ (Table 2c), allowing the urban tile to store a large fraction of Q^* . The radiative coupling to the ground means that the results can also be affected by differences in soil-temperature states between configurations. In London, for example, the HAD-B soil in each layer is colder by $\sim 2 \text{ K}$ compared with CTRL-B (Online Resource Fig. S8a), which may lead to a reduction of the surface temperature and hence L^\uparrow . Soil temperatures would be reduced further if HAD-B used the region-T land cover (not shown), leading to an improvement in grid box L^\uparrow in MAM and JJA as f_{urban} is lowered (Online Resource Fig. S4b). However, the over-prediction in spring/summer persists if Q_F is accounted for in HAD-B, as illustrated by the CTRL-B results (Fig. 5b).

In MORUSES, the different heat capacities for canyon and roof (C_c , C_f ; Table 2c) allow large differences in thermal inertia. The roof areal heat capacity is computed from $C_f = c_f \Delta z_f$. In all MORUSES runs, this is equal to the UKV-M value of $\sim 0.04 \text{ MJ K}^{-1} \text{ m}^{-2}$, resulting from the ‘thin roof’ setting with $\Delta z_f = 0.02 \text{ m}$, which is

an order of magnitude lower than the computed material thickness of walls and road ($\Delta z_{r,w} \propto k_{r,w}^{0.5}$). The canyon areal heat capacity, as the sum of wall and road components, is directly proportional to the height-to-width ratio: $C_c = 2HW^{-1}c_w \Delta z_w + c_r \Delta z_r$ (Porson et al. 2010a). The large effect of H/W on C_c is illustrated in the values of the four London/KCL source areas (Fig 4). C_c varies between 0.41 (region C; $H/W = 0.36$) and $1.18 \text{ MJ K}^{-1} \text{ m}^{-2}$ (region A; $H/W = 1.66$; Table 3a, c). For region T, it takes a value of $0.9 \text{ MJ K}^{-1} \text{ m}^{-2}$ ($H/W = 1.18$; Table 2a, c). The heat-storage capacity of the MORUSES canyon in London is larger than for the Best-1T urban slab (Table 2c). An approximate bulk C_U for MORUSES can be computed from C_c and C_f based on the proportions of roof and canyon, resulting in $C_U = 0.42 \text{ MJ K}^{-1} \text{ m}^{-2}$ for CTRL-M (region T), followed by UKV-M (west) and UKV-M (east) with 0.33 and $0.26 \text{ MJ K}^{-1} \text{ m}^{-2}$, respectively (Table 2c).

As the areal heat capacity increases, the diurnal amplitude of T_* decreases. As a result, L^\uparrow decreases during the day as more heat is stored in the urban fabric, and increases at night when it is released back into the atmosphere. The comparatively large areal heat capacities of the Best-1T slab and the MORUSES canyon, together with the radiative/conductive coupling to the soil, result in a phase delay of L^\uparrow . However, in MORUSES, this lag is partially compensated by the fast response of the insulated roof to radiative forcing and its contribution to the weighted average of L^\uparrow for the grid box, which contributes to the improved agreement with the observations. The effects of H/W and W/R on MORUSES are further discussed in Section 6.3.

The good match between modelled and observed Q^* for the MORUSES runs and CTRL-B can be explained partially by compensating errors between K^\uparrow and L^\uparrow (Figs. 5c, 6c) and the large role of the observed K^\downarrow (model forcing) in Q^* . The London HAD-B run has the poorest 24-h MAE, and under-predicts Q^* in spring and summer as a result of the over-prediction of both outgoing radiation components (Fig. 5c).

6.1.2 Turbulent heat fluxes

Modelled turbulent heat fluxes (grid box-averaged) agree better with the observations in suburban Swindon (Fig. 6d,e) compared with central London (Fig. 5d,e), as was found in the SUEWS land-surface model evaluation at these sites (Ward et al. 2016). For both Q_H and Q_E , the configuration variability is lower in Swindon as parameter and physics differences of the urban schemes are partially masked by the larger contribution of non-urban surfaces compared with London (Table 1).

In London, Q_H from CTRL-B and HAD-B has a consistent phase shift of up to ~ 3 h relative to the observations (Fig. 5d). Differences between the two Best-1T runs can be attributed to (i) including realistic Q_F emissions in CTRL-B rather than assuming $Q_F = 0 \text{ W m}^{-2}$ as in HAD-B; (ii) the lower bulk albedo in CTRL-B resulting in a larger Q^* ; and (iii) the increased contributions of non-urban tile fluxes to the grid box average in CTRL-B (Table 1a), somewhat reducing the phase delay from the urban tile. The MORUSES runs (CTRL-M, UKV-M) capture the phase of Q_H better since the large differences between the canyon and roof SEBs are represented. The remaining delay in the timing of the grid box Q_H for the two UKV-M runs can be attributed to the larger W/R (Tables 2a, 3a) compared with the sub-sampled CTRL-M results, which are dominated by region A (52.6%) and D (32%). The small MORUSES C_f results in a faster and earlier increase of the roof T_* compared with the canyon, leading to an earlier phase and increased amplitude of Q_H . The larger thermal inertia of the canyon reduces the diurnal variability of the canyon temperature and delays the daytime Q_H peak.

For the London site, the largest relative differences of the peak Q_H between configurations occur in SON and DJF (Fig. 5d), when anthropogenic heat emissions from space heating are higher. The observed SON/DJF Q_H does not become negative at night, and is higher in winter than autumn (median of ~ 50 cf. $\sim 40 \text{ W m}^{-2}$). The DJF Q_H for HAD-B, which neglects Q_F , is negative over large parts of the day. Using JULES- Q_F (Fig. 2) in UKV-M improves this, but the results are still closer to HAD-B than to the CTRL runs that use a higher Q_F representative of central London from GQF. In winter, the monthly median GQF Q_F (Fig. 2) is larger than the observed median peak Q^* (Fig. 5c), and hence makes a large contribution to the energy balance. The impact of the representation of Q_F on the modelled Q_H is discussed in Section 6.2. Adjusting Q_F supplied to the CTRL runs by the local urban fraction (cf. Section 3.6), so that it is consistent with the way in which JULES applies it, increases the magnitude of the grid box Q_H , which in London leads to better daytime results in winter and autumn, but a notable over-prediction at night in all seasons (Online Resource Fig. S1d). The positive nocturnal bias is a result of not using the diurnal cycle of the GQF results, which have much higher Q_F during the day. The smaller f_{urban} in the Swindon grid box makes the Q_F -scaling effect less prominent (Online Resource Fig. S2d).

In Swindon, Q_F contributes less to the energy balance and the seasonal variability of model performance is reduced compared to London. HAD-B is the only model configuration consistently underestimating Q_H . However, if HAD-B has the CTRL land cover, the performance is

similar to CTRL-B (Online Resource Fig. S5d), i.e. the difference can be attributed to the ~ 10 -km HadGEM3-PRIMAVERA ancillaries with more non-urban surfaces assumed. Hence, compared with CTRL-M/B ($f_{\text{urban}} = 0.49$) and UKV-M (0.31; Table 1b), HAD-B has larger Q_E (Fig. 6e). The JULES-ULSMs over-predict the sensible heat flux at night and do not reproduce the negative nighttime Q_H in Swindon in summer (Fig. 6d), similar to SUEWS (Ward et al. 2016).

At both sites, the ratio of roughness lengths of heat and momentum computed by MORUSES is larger than the default Best 1T value of 10^{-7} , in the case of CTRL-M at Swindon by almost five orders of magnitude (Table 4), leading to a larger amplitude of Q_H (cf. Fig. 3 in Best et al. 2006). Variability of the roughness ratio among the MORUSES runs is related to differences in local morphology.

Large seasonal variations of modelled Q_E in London (Fig. 5e) are related to the presence of the River Thames and urban vegetation (Fig. 4). The CTRL and UKV-M (east) runs have considerable water-tile fractions (Table 1a). This is not the case for UKV-M (west) and HAD-B, for which $f_{\text{urban}} = 0.95$ and 0.98 , and Q_E is strongly underestimated during the day. For HAD-B this results in near-zero median Q_E in all seasons as only rainfall is available for evaporation, but removed quickly from the impervious surface (Section 6.4). Including the river and vegetation in HAD-B using the region T land cover would greatly improve Q_E (Online Resource Fig. S4e), while in the absence of anthropogenic heat sources Q_H would be less well predicted (Online Resource Fig. S4d).

Urban-tile evaporation in Best 1T and MORUSES could be enhanced by allowing for some pervious areas, but implications (e.g. regarding soil-moisture transport) are currently unclear, as are implications of prescribing urban irrigation or other anthropogenic moisture sources.

Observations indicate that KCL is not as strongly affected by evaporation from the river as its proximity suggests (Kotthaus and Grimmond 2014b). This is because of the river surface being vertically displaced below the local ground level and hence the effective height of the flux tower being greater relative to the river. Kotthaus and Grimmond (2014b) discuss that for advection over the river ($100^\circ \leq \theta \leq 230^\circ$), with a tidal induced variation of the water surface of ~ 6 m, the eddy-covariance sensor heights can be above the river's growing internal boundary layer. This is most notable for wind directions at right angles to the river, i.e. region C (Fig. 4), whereas for flow from regions B and D evaporation from the river is more likely measured. This feature of the observations was also identified as a main factor contributing to the positive Q_E bias in SUEWS at KCL (Ward et al. 2016).

The source-area sub-sampling of the CTRL runs results in the grid box Q_E being mainly affected by the lake tile in summer, leading to a positive offset of the median diurnal cycle and a much larger 75th percentile compared with the observations. Otherwise, the CTRL runs are dominated by evapotranspiration from broadleaf trees (Table 1a). Since the UKV LAI is 0 for all PFTs (Section 3.5), JULES treats the vegetated tiles as bare soil, for which Q_E shows a pronounced daily cycle. The river Q_E has a weaker diurnal amplitude. The lower water fraction in UKV-M (west) results in the negative bias relative to UKV-M (east).

The LAI anomalies (Section 3.5) in Swindon affect the tree and shrub tiles in UKV-M, but not grass, which has the highest plan-area fraction (49%; Table 1b). At this site, the CTRL runs with the lowest vegetation fraction have the best agreement of modelled Q_E with the observations, but all configurations overestimate the latent heat flux in spring and summer (Fig. 6e). The fact that HAD-B in Swindon has the wettest soil (Online Resource Fig. S8b) appears to be a secondary effect to having the largest vegetation fraction when it comes to the over-estimation of the grid box Q_E . Analysis of the grass-tile Q_E shows that HAD-B, although it behaves similarly to the much drier UKV-M, has larger amplitudes of Q_E in summer and autumn (not shown). Prescribing land-cover representative of the source area of the observations in HAD-B in Swindon reduces Q_E and increases Q_H , resulting in a skill comparable with CTRL-B (Online Resource Fig. S5d, e). However, climate applications are clearly dependent on appropriate ancillaries.

6.1.3 Diurnal variations of model performance

Time-of-day-dependent model performance (Section 5.2) for the turbulent heat fluxes is evaluated by season using the MAE (Eq. 4; Fig. 7). Corresponding MBE (Eq. 5) metrics are given in Fig. S6 of the Online Resource. At both sites, the largest cross-configuration variability occurs in the morning and afternoon. In Swindon, for all models, the afternoon MAE (Fig. 7b) is larger than for other periods of the day. Here, HAD-B has large |MBE| (Fig. S6b) for both Q_H and Q_E in the morning and afternoon. In spring and summer in London, all models except for CTRL-M have a large negative Q_H bias in the morning (Fig. S6a). This is associated with the characteristic phase delay of the Best-1T model and the lags determined in the diurnal cycles of the UKV-M runs (Fig. 5d). Compared with suburban Swindon, the central London Q_E bias is quantitatively more similar for different time periods (Fig. S6a).

The model errors for the two UKV-M configurations in London (east/west grids; Fig. 4) indicate that neither grid characteristics provide a consistently good representation of the actual source-area properties of the observations.

The morning and afternoon MAE are reduced for UKV-M (west) as a result of the increased urban tile fraction (Fig. 7a), but the model tends to over-predict the afternoon Q_H in summer (Fig. S6a). Q_E , however, has the lowest errors for UKV-M (east) of all the configurations considered.

The transition periods around sunrise and sunset not exhibiting large absolute errors implies that the models respond accurately to changes in radiative forcing. However, relative errors may be large when the fluxes are small.

6.1.4 Performance summary

Taylor diagrams (Taylor 2001) summarise the model performance at London/KCL (Fig. 8a) and Swindon (Fig. 8b). Metrics shown for the entire evaluation period at both sites include the correlation coefficient (Eq. 7), the normalised standard deviation (σ_P/σ_O), which is 1 for a perfect prediction, and the normalised centred root-mean-square error (nRMSE; Eq. 6; 0 for a perfect prediction).

For all energy-balance components and model configurations, predictions are poorer at the high-rise, high-density site in London, compared with suburban Swindon. At both sites, errors are larger for the turbulent heat fluxes than for the radiation components, with Q_E showing the largest cross-configuration variability, particularly in London (Fig. 8a), which can be attributed to the geometric complexity of the site and the variations of the prescribed land-cover fractions between configurations (Table 1a). Even after using high-resolution morphology ancillaries, more realistic monthly Q_F values and source-area sub-sampling, Q_H and Q_E in the CTRL-M/B runs still have large uncertainties in central London.

While the HAD-B performance is better for Swindon (Fig. 8b), model errors are consistently larger compared with the other configurations. Comparisons with CTRL-B show that this is in part from the HAD-B land-cover ancillaries. Model skill can also be improved by using suitable bulk parameters (e.g. α_U) and, particularly in London, accounting for anthropogenic heat emissions. While the phase lag of the Best-1T scheme at the suburban site is masked by larger non-urban contributions to the grid box Q_H , the delay strongly affects the performance in highly urbanised environments such as central London.

6.2 Contribution of Q_F

The seasonal diurnal cycle of Q_H in London (Fig. 5d) highlights the large impact of anthropogenic heat emissions on the local USEB when contrasting the CTRL runs with UKV-M and HAD-B. To explore the effect within the currently operational JULES-UKV framework, UKV-M (east) is repeated with (i) the Q_F source switched off (as

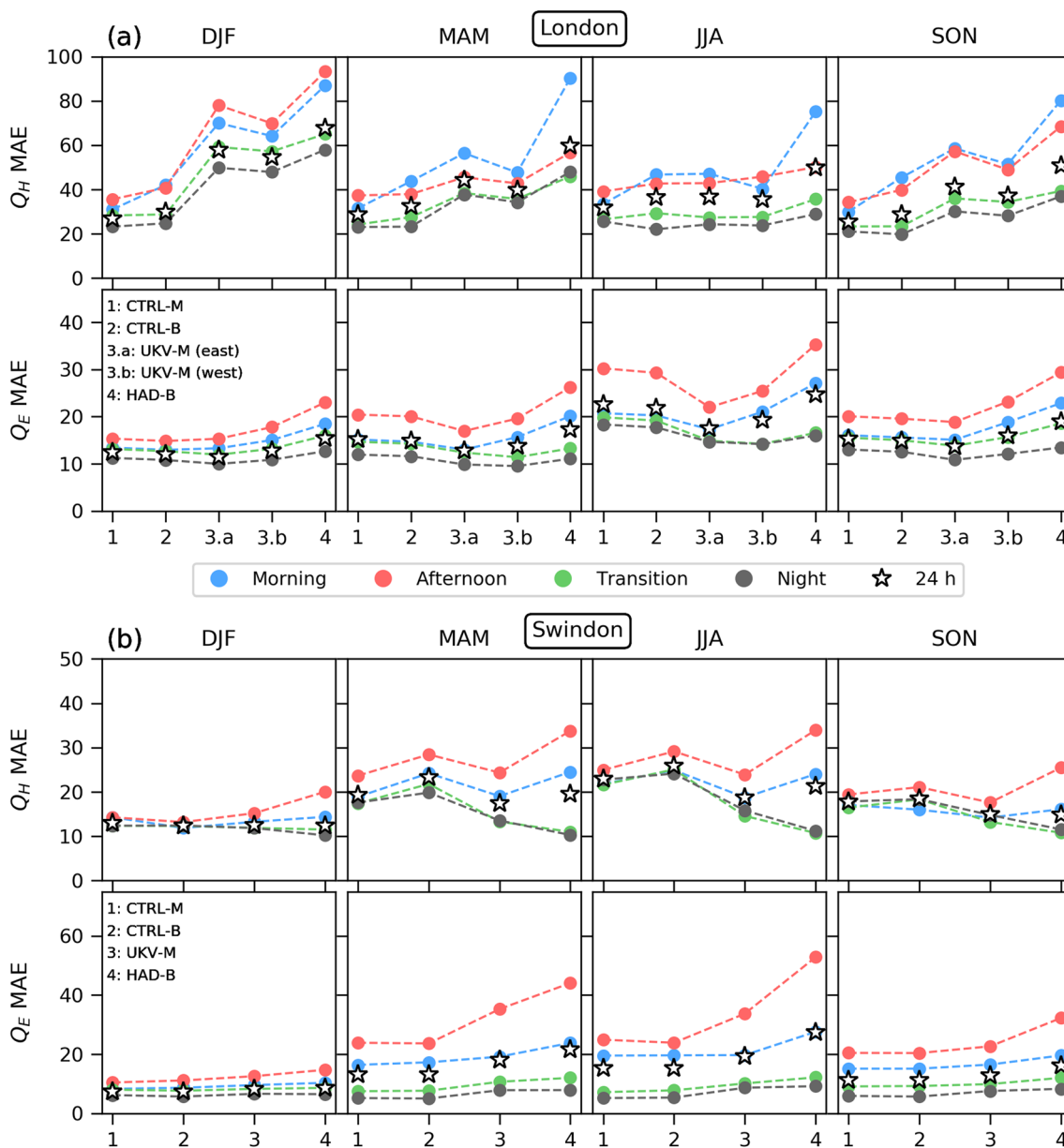


Fig. 7 MAE ($W m^{-2}$) of Q_H and Q_E for (a) London/KCL and (b) Swindon per season and by time-of-day (Section 5.2) for each model configuration. Dashed lines are used for ease of visualising relations between model configurations. 24-h values (stars) are given in Figs. 5 and 6

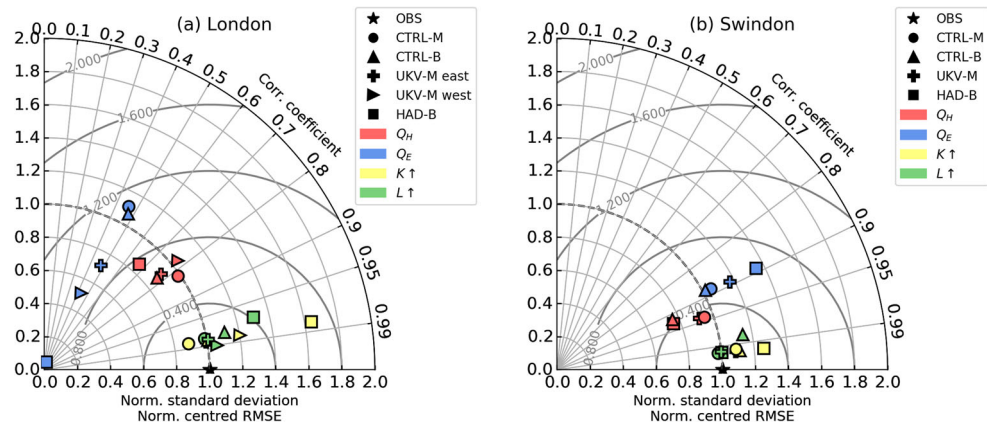
in HAD-B) and (ii) using monthly median (unscaled) GQF data (Fig. 2).

Switching off Q_F entirely only slightly decreases the Q_H model skill compared with the default run using monthly JULES- Q_F values that are currently prescribed across the UK (Fig. 9). While the quantitative differences are small, qualitatively the effect of using JULES- Q_F is not negligible as it leads to $Q_H < 0 W m^{-2}$ occurring less frequently in autumn and winter, which in turn affects temperature profiles above the city. Prescribing the GQF values distinctly improves the results (UKV-M GQF;

Fig. 9), with the 24-h MAE and MBE values now being comparable with CTRL-M (Fig. 5d).

Diurnal and weekday/weekend variations of heat emissions have been removed from the prescribed GQF data by following the current JULES approach of using only monthly values (Section 3.6). At KCL, averaging the hourly GQF output results in smaller (greater) daytime (night-time) Q_F values. This could explain some of the remaining differences to the observations in autumn and winter (UKV-M GQF, Fig. 9; CTRL-M, Fig. 5d), when Q_F is both greatest and shows the largest temporal variability (Fig. 2).

Fig. 8 Taylor diagram for (a) London/KCL (3-year evaluation period) and (b) Swindon (2-year period) based on normalised standard deviation (σ_P/σ_O ; thin arcs), normalised centred RMSE (thick arcs) and the correlation coefficient



6.3 MORUSES morphology controls

In the operational JULES–UKV, the empirical relations (Bohnenstengel et al. 2011; Gilham et al. 2019) used to obtain H , H/W and W/R as a function of f_{urban} (Section 3.2) reduce the possible morphology variations from three degrees of freedom to one as the parameters are linked. The Section 6.1 MORUSES results demonstrate the influence of street-canyon morphology (H/W) and roof-canyon partitioning on the grid box-averaged fluxes. While canyon geometry affects the bulk canyon parameters (e.g. α_c , C_c), W/R determines the contributions of canyon and roof to the grid box average and, via the Macdonald roughness parameters, the value of z_h .

Figure 10 shows Q_H and thermal inertia, $C dT_*/dt$, modelled with MORUSES in response to varying (i) W/R ($H/W = 1.18$ fixed; Fig. 10a–c) and (ii) H/W ($W/R = 0.44$ fixed; Fig. 10d–f). The mean building height ($H = 19.81$ m) is held constant, which conceptually implies that for (i) both H and W remain fixed while the roof length changes, whereas for (ii) both canyon and roof lengths change so that both H and W/R remain constant.

The sensitivity test uses UKV–M (London/KCL) for a completely urbanised grid box ($f_{\text{urban}} = 1$). The fixed parameter values of H , H/W and W/R are the CTRL values for region T (Fig. 4; Table 2a), while thermal and radiative parameters are the JULES–UKV defaults (Table 2b, c). Results (Fig. 10) are shown for the median JJA diurnal cycles, but are qualitatively similar throughout the year.

In the sensitivity test, MORUSES computes z_0 and z_d with the Macdonald method (Section 3.3). While z_0 depends on H , H/W and W/R , z_d only varies with H and W/R and hence remains constant if H/W is varied (Eqs. 8,9). The MORUSES resistance network depends on H/W , z_0 and z_d . Therefore, the resultant roughness lengths for heat (z_h ; Eq. 10) of the canyon and roof tiles change throughout the analysis. For the results in Fig. 10, the corresponding curves for z_0 , z_d and z_h for varying W/R (Fig. 12a) and H/W (Fig. 12b) are discussed in Appendix 2.

The contrast between roof and canyon tile fluxes is drastic. With the very small heat capacity of the roof (Table 2c), the corresponding tile Q_H has much larger amplitudes and peaks earlier in the day (similar in timing

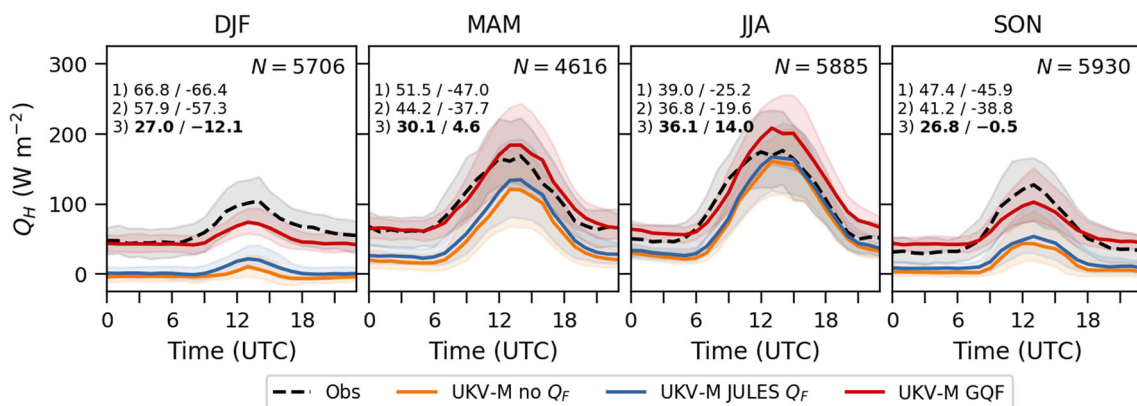


Fig. 9 As Fig. 5d, but for UKV–M (east) with different prescribed monthly Q_F values: 0 W m⁻² (no Q_F), default JULES– Q_F (results as in Fig. 5d) and monthly median (unscaled) GQF values (cf. Fig. 2)

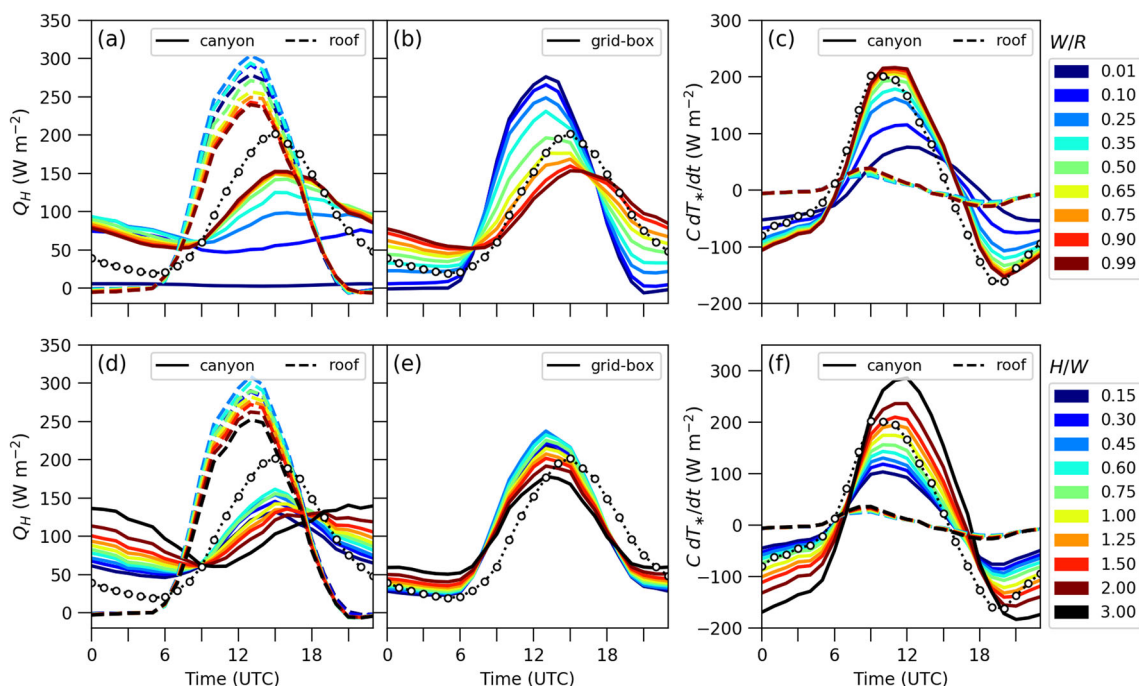


Fig. 10 Response of the median summer (JJA) diurnal cycles of UKV-M Q_H and $C dT_*/dt$ to changes in MORUSES morphology parameters: (a–c) canyon fraction (W/R) and (d–f) height-to-width ratio (H/W), setting f_{urban} to 1 at London/KCL. In all cases, $H =$

19.81 m is held constant. In (a–c) H/W is fixed at 1.18, in (d–f) W/R is 0.44 (CTRL region T values; Table 2a). The respective Best-1T reference curves for the urban tile (dotted lines with circles) are shown as a reference using CTRL-B parameters (Table 2)

to Q^* ; Fig. 5c) compared with the canyon (Fig. 10a,d), but contributes only marginally to the net storage heat flux (Fig. 10c, f).

For a very small canyon fraction ($W/R = 0.01$; Fig. 10a), the canyon z_h defaults to 10^{-30} m in this test (Fig. 12a), leading to a very low canyon Q_H (Fig. 10a). The roughness length defaults to $z_{0m} = 0.05$ m for $W/R \leq 0.25$, while the roof z_h increases before it continuously decreases from a maximum at $W/R = 0.25$ (Fig. 12a), which leads to the initially increasing then decreasing peaks of the roof Q_H (Fig. 10a). Once the canyon fraction is larger than ~ 0.5 , Q_H and thermal inertia vary very little.

Increasing W/R successively reduces the diurnal amplitude of the grid box Q_H and delays the timing of the peak values up to ~ 3 h (Fig. 10b) as the contribution of the roof tile to the plan-area averaged fluxes decreases. For small canyon fractions, the grid box results reflect the quick response of the (insulated) MORUSES roof to radiative forcing, resulting in very high/low roof surface temperatures during the day/night. The larger thermal inertia of the canyon (Fig. 10c) means smaller diurnal variations of the canyon surface temperature and hence smaller temperature gradients between canyon and atmosphere. In all test scenarios, the nocturnal roof Q_H is negative (Fig. 10a,d). For W/R between 0.1 and 0.9, there is a ~ 100 $W m^{-2}$ difference in the peak values of the grid box Q_H and the diurnal amplitude decreases from ~ 275 to ~ 100 $W m^{-2}$ (Fig. 10b)

as the heat-storage amplitude increases (Fig. 10c). As the canyon fraction increases, MORUSES approaches the Best-1T modelling framework of a single impervious slab with large thermal inertia and the weighted ratio of canyon and roof z_h to z_0 approaches the Best-1T value of 10^{-7} (Fig. 12a).

Varying H/W at a fixed $W/R = 0.44$ in the isolated-building regime ($H/W \leq 1/3$) results in the model defaulting to $z_h = 0.1z_{0h}$ for canyon and roof (only at $H/W = 0.3$; Fig. 12b), which contributes to the non-monotonic behaviour evident in the amplitudes of the tile and grid box Q_H (Fig. 10d, e). The variation of the grid box Q_H (Fig. 10e) between $H/W = 0.3$ (isolated-building regime) and 3.0 (skimming flow) is smaller compared with the response to changes in W/R , with differences in the daytime peaks of ~ 60 $W m^{-2}$ and diurnal amplitudes reducing by ~ 100 $W m^{-2}$.

The large effect of H/W on the areal heat capacity of the canyon and therefore on its thermal inertia (Fig. 10f) affects both the amplitude and phase of the canyon Q_H (Fig. 10d). As more heat can be stored with increasing H/W , temperature gradients become smaller during the day and larger at night when the heat is released ($C_c dT_*/dt < 0$). For $H/W = 3.0$, which is a typical value for high-rise city centres, the storage heat flux into the canyon has become very large during the day (~ 290 $W m^{-2}$; Fig. 10f) so that Q_H only starts to rise around noon and reaches

its maximum at midnight (Fig. 10d). A similar pattern is present in the canyon Q_H for region A of the CTRL–M run at KCL ($H/W = 1.66$; $W/R = 0.41$; Table 3; not shown). The night-time canyon Q_H increases by approximately a factor of 2 when increasing H/W from 0.3 to 3 (Fig. 10d). Although the presence of the roof dampens this response, the grid box-averaged Q_H is still increased noticeably at night (Fig. 10e).

The analysis shows that the availability of accurate urban morphology ancillaries is crucial if MORUSES is to represent the urban energy-balance successfully. Constraining the proportion of the canyon, which can be more easily derived from satellite data, has a larger impact on the results than H/W . Repeating the sensitivity test by allowing H to vary to obtain different W/R or H/W leads to quantitatively similar results (not shown).

6.4 Meteorological controls

The hourly model bias ($P_i - O_i$) of the turbulent heat fluxes and L^\uparrow from UKV–M in London (east grid) and Swindon is analysed as a function of meteorological forcing, focusing on air temperature (T_{air} ; Fig. 11a, b), horizontal wind speed (U_h ; Fig. 11d) and rainfall (hours after rain; Fig. 11e, f). Evaluating the bias with horizontal wind direction (θ , Fig. 11c), unaccounted for in the JULES tiling approach, is useful to assess the impact of subgrid-scale land-cover heterogeneity.

Atmospheric stability is indicated using the stability parameter $\zeta = (z - z_d)/L$, where L is the observed Obukhov length and $z - z_d$ is the effective height of the observations, equivalent to the JULES forcing height (Section 4). Stability bins are selected as follows: $|\zeta| \leq 0.05$ (near neutral), $\zeta < -0.05$ (unstable), $\zeta > 0.05$ (stable). Over the evaluation periods, stable atmospheric stratification occurs more often in suburban Swindon (15.6%) than in central London (0.9%). In London, the observed prevalent stability condition is unstable (74% versus 16.3% in Swindon), while in Swindon near-neutral conditions prevail (68.1% versus 25.1% in London).

The bias of Q_H driven by T_{air} (Fig. 11a) differs between the two sites. The increasing negative model bias with decreasing T_{air} in central London supports the earlier assessment of the large importance of Q_F . Once T_{air} is greater than $\sim 18^\circ\text{C}$, the median Q_H bias is strongly reduced, while the overall scatter increases. For Swindon, where Q_F is much lower (Fig. 2), the median bias varies little over a wide range of forcing temperatures. At both sites, the statistical uncertainty of the binned bias at relatively high air temperatures ($T_{\text{air}} > 25^\circ\text{C}$) is affected by reduced sample sizes ($100 > N \geq 20$).

The variation of the L^\uparrow bias with T_{air} (Fig. 11b) at both sites indicates that the surface is warmer in the observations

compared with the model. In Swindon, the bias is relatively uniform across the range of forcing temperatures once $T_{\text{air}} > 0^\circ\text{C}$, while in London the magnitude of negative bias increases for $T_{\text{air}} > 20^\circ\text{C}$. At KCL, the radiometer source area contains mostly buildings and roads, while in Swindon L^\uparrow is affected both by impervious and vegetated surfaces (Ward et al. 2013). This could indicate that the model bias is of systematic nature.

As expected, given the fairly homogeneous land-cover distribution around the Swindon site (Ward et al. 2013), the Q_H bias shows very little variation with wind direction (Fig. 11c). In London, however, there is a distinct response to the patchiness of the distribution of buildings, water and vegetation surrounding the site (Fig. 4). The reduction of the negative median Q_H bias as the flow approaches the flux tower from the river ($100^\circ \leq \theta \leq 230^\circ$) can result from (i) the increased importance of the observed Q_E relative to Q_H and (ii) decreasing contributions of Q_F . This indicates that the UKV–M (east) results (Fig. 5) for the given tile distributions (Table 1a) represent the site characteristics best for southerly wind directions, when the observations are less affected by Q_F and contain signatures of the river, while a fetch over the urbanised north is associated with too low Q_H .

Evaluating the Q_H bias as a function of wind speed (Fig. 11d) is useful to identify the relative importance of larger scale horizontal advection of heat and moisture. While advection affects the observed turbulent heat fluxes, it cannot be modelled offline. It is noted that the different wind-speed characteristics in London and Swindon are affected by the measurement height of the observations (Section 4). At both sites, the variation of the median Q_H with U_h is low over the range of wind speeds observed, while the bias range (between 5th and 95th percentiles) increases after U_h exceeds $\sim 4 \text{ m s}^{-1}$. This is also the case for Q_E (not shown). This could indicate that for high wind speeds, advection can become important especially if the offline results are evaluated over shorter (non-climatological) periods. Furthermore, under neutral stability conditions (high wind speeds), the flux footprint becomes much larger (Fig. 7 in Kotthaus and Grimmond 2014b), which introduces changes in the observed heat-flux characteristics.

In London, high wind speeds are more likely to occur from the dominant wind-direction sector of $180\text{--}225^\circ$ (covering region B; wind rose in Fig. 4), which is associated with a larger positive Q_H bias in UKV–M (east) (Fig. 11c) as evaporation from the river is more likely to affect KCL for flow from this direction. The larger bias variability for high wind speeds at KCL (Fig. 11d), in particular the tendency towards an over-prediction of Q_H (increased 95th percentiles), is likely impacted by these changes in energy partitioning with θ .

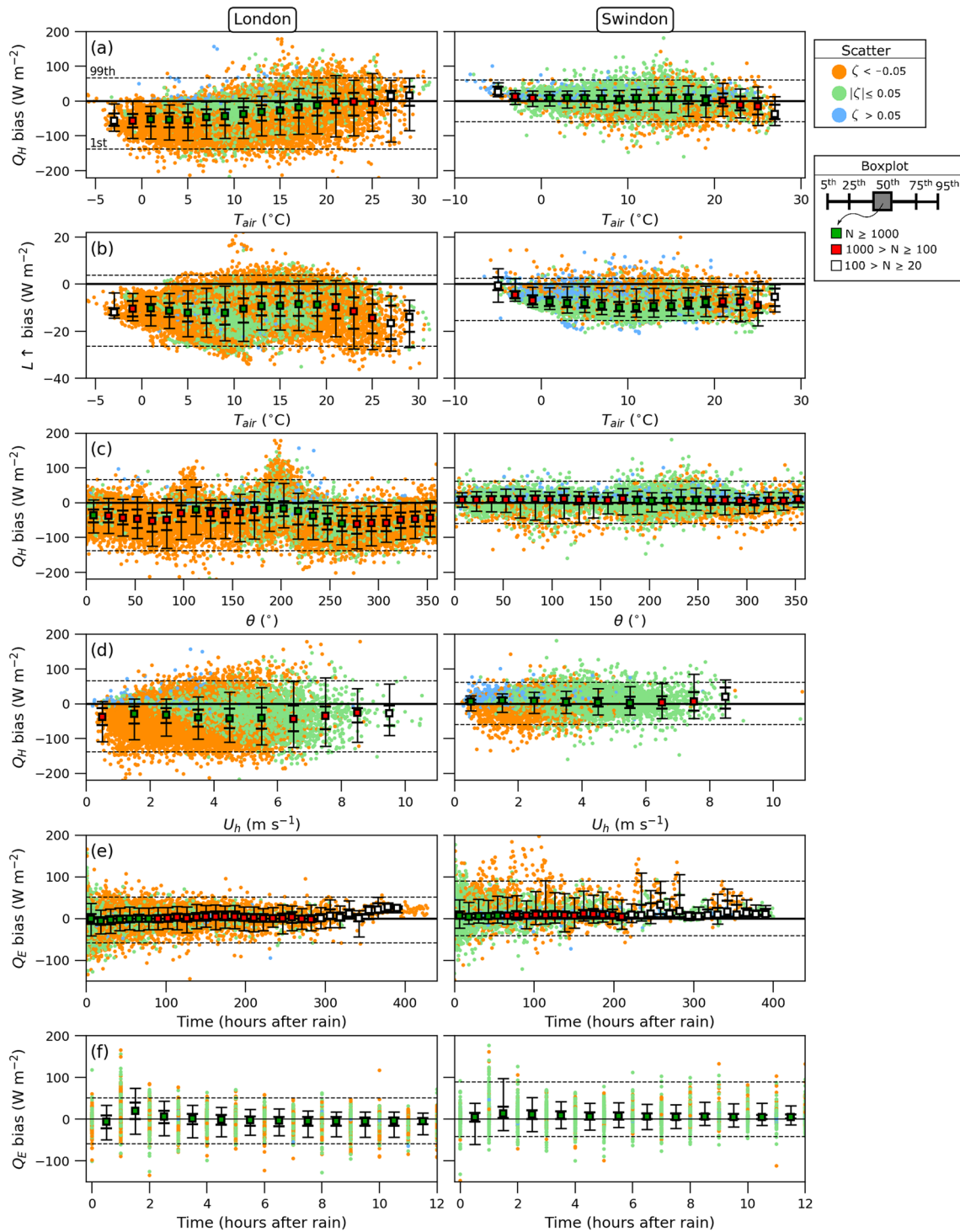


Fig. 11 Hourly UKV-M model bias ($P_i - O_i$; scatter coloured by stability parameter $\zeta = (z - z_d)/L$) of Q_H for 3 years at London/KCL (east grid, Fig 4) and 2 years in Swindon in response to (a) air temperature (T_{air}), (c) wind direction θ , (d) wind speed (U_h), (b) model bias of $L \uparrow$ as a function of T_{air} and (e, f) model bias of Q_E depending on

hours after rainfall with (f) showing the bias over the initial 12 h. The 1st and 99th percentiles of the hourly biases are shown (dashed lines). Boxplots are for bins of (a, b) 2°C, (c) 15°, (d) 1 m s⁻¹, (e) 12 h and (f) 1 h

Wind speed also affects Q_H via the MORUSES resistance network. The resistance to scalar transport is inversely proportional to the modelled canyon velocities

(Harman et al. 2004). These scale with the velocity at the top of the canyon that is derived by interpolation from the forcing wind speed. Higher wind speeds imply

lower resistances and thus higher turbulent fluxes. The interpolation to the canyon top critically depends on z_d and z_0 , which MORUSES estimates from the Macdonald method that can underestimate the surface roughness in heterogeneous building environments (cf. Section 3.3).

To determine the model response to precipitation, the Q_E bias is evaluated as a function of hours after the last rainfall event (Fig. 11e). For short periods after rain, evaporation from the impervious urban tiles can be expected to make the largest relative contribution to the overall latent heat flux. In London, where impervious surfaces contribute with 79% to the prescribed land cover (UKV–M east; Table 1a), the variability over the first 12 h following rain is large and followed by a negative median bias over the next 12-h interval, indicating that the model represents conditions as drier than observed. This feature is not as clear in Swindon, where f_{urban} is only 0.31 (Table 1b) and the vegetation contribution to Q_E is dominant. When evaluating only the first 12 h after rain (Fig. 11f), at both sites there is an increased positive model bias of Q_E over the first 1 to 2 h, which could indicate that the urban tiles dry out too quickly.

JULES partitions precipitation into throughfall, interception, runoff and infiltration (Essery et al. 2001). On the urban tile(s), precipitation reaches the surface without interception and the partitioning of rainwater into infiltration and runoff strongly favours the latter. Surface infiltration is curbed by using a low (0.1) attenuation factor for the saturated hydraulic conductivity, which together with the low value of prescribed surface water capacity (0.5 kg m^{-2}) results in relatively large urban runoff (Eq. 48 in Best et al. 2011). Runoff characteristics of the JULES–UKV (CTRL–M/B, UKV–M) and JULES–GL7.0 (HAD–B) are affected by the use of different models to represent soil-moisture heterogeneity (Appendix 1). While JULES–UKV uses PDM, JULES–GL7.0 uses TOPMODEL (for details see Best et al. 2011), which treat the saturation-excess runoff differently. At London/KCL, all configurations generate large runoff rates within a single model time step, which strongly limits evaporation (Online Resource Fig. S7a). The increased vegetation fraction in Swindon reduces the grid box runoff as rain can be intercepted and held by the vegetation canopy (Online Resource Fig. S7b). At both sites, runoff is generated for precipitation rates exceeding $\sim 0.9\text{--}1.2 \text{ mm h}^{-1}$ (JULES–UKV) or $\sim 1.3\text{--}1.7 \text{ mm h}^{-1}$ (JULES–GL7.0) based on a linear fit of the data.

7 Conclusions

The Best-1T scheme and the 2-tile canopy model MORUSES in JULES are evaluated using multi-year flux observations in the highly urbanised city centre of London and in suburban Swindon, UK. The model performance

is studied under different forcing conditions and using different resolution ancillary information (land cover, building morphology). Simulations based on the JULES science configurations and ancillaries used in the UM for regional NWP (JULES–UKV; 1.5 km resolution) and global high-resolution ($\sim 10 \text{ km}$) climate simulations (HadGEM; JULES–GL7.0) are compared with control runs that use model parameters derived from observations and high-resolution ($\sim 1 \text{ m}$) GIS data and more representative Q_F . With respect to the study aims, the following conclusions are drawn:

1. Implications of the JULES–ULSM options and resolution/accuracy of ancillaries

- JULES with MORUSES and Best 1T performs better at the suburban site, where results to a larger degree are determined by the vegetation modelling, compared with the high-rise, high-density site in central London, where physics differences between the ULSMs are more pronounced.
- At both sites, the outgoing short- and longwave radiation are better represented than the turbulent heat fluxes. The cross-configuration variability of Q_H and Q_E is particularly large in London because of the site complexity (and its representation in the ancillaries) and the different ways in which the ULSMs represent the diurnal variability of surface temperatures.
- For both MORUSES and Best 1T, the prediction of Q_H in London is poorest in winter and autumn (increasingly negative median bias in UKV–M below $T_{\text{air}} \approx 18^\circ\text{C}$), when the estimated contribution of Q_F to the local energy balance is of a similar magnitude as Q^* . The seasonal performance variability is smaller in Swindon, where Q_F is low year-round. Using a more representative Q_F in the CTRL runs for central London improves the model skill.
- Best 1T with the climate-model land cover and default JULES–GL7.0 urban parameters at both sites shows the largest difference to the observations. The HadGEM3–PRIMAVERA land-cover ancillaries overestimate f_{urban} in London, resulting in near-zero median Q_E , while underestimating urban land cover in Swindon (too high Q_E) at the $\sim 10\text{-km}$ grid scale. The relatively high default bulk urban albedo (0.18) and not accounting for Q_F further contribute to differences from the observed SEB. The characteristic phase delay of Q_H (negative bias in the morning) and the positive afternoon bias of L^\uparrow are more apparent in environments with high f_{urban} like London.
- The presence of the roof in MORUSES and its fast response to radiative forcing improves the timing of Q_H . The high thermal inertia of the MORUSES canyon means that the phase and amplitude of the (weighted)

grid box Q_H critically depends on the prescribed plan-area fractions of roof and canyon tiles, which in the UKV are parameterised through f_{urban} . Variations of canyon proportion (W/R) have a larger effect on the amplitude of Q_H than variations of H/W .

- MORUSES and Best 1T both limit the water-holding capacity on the urban tile(s) so that rainwater is drained too quickly and the tiles dry out fast. This subsequently limits evaporation and, if no pervious surfaces occur in the grid box (e.g. vegetation or water), can result in errors in the energy partitioning.

2. Potential for model improvements to support model outputs being used for integrated urban services

- Anthropogenic forcing: Providing a more realistic and flexible way of representing anthropogenic heat emissions is essential to improve the performance of the JULES ULSMs, particularly for applications in high-density city centres. This critically needs to include spatial and temporal variations of Q_F either through ancillaries or (better) through modelling and should allow for the representation of the interplay between human activity and the urban micro-climate (e.g. feedback between the use of space heating/cooling and outdoor air temperatures). Such options are currently being explored. Similarly, anthropogenic emissions of water (e.g. through street cleaning or irrigation of urban vegetation) can be important in some cities (e.g. Ao et al. 2018; Dou et al. 2019), but are currently not modelled in JULES.
- Land-cover information: Providing accurate land-cover information is crucial for the model performance. In many cities, urban vegetation (street trees, parks) or water bodies make a non-negligible contribution to the local urban climate. Not representing the sub-grid scale land-cover variability in cities in the model ancillaries can lead to an under-representation of the urban Q_E . In such cases, applications that rely on heat-flux ratios (e.g. Bowen ratio; evaporative fraction) are strongly impacted, which can affect estimates of water requirements for urban vegetation (e.g. irrigation planning) or health-related applications (e.g. heat stress; thermal comfort).
- Morphology information: The operational UKV MORUSES morphology parameterisations were derived for London. Hence, access to high-resolution building morphology data is necessary if the model is used elsewhere. For UK applications, it also has to be considered that the empirical relations were derived for ~ 1 -km NWP applications, and it is unclear whether they are still appropriate at much higher or lower resolutions. Over London, the UKV is currently run routinely at ~ 300 -m horizontal resolution, and with even higher

resolved simulations being used (~ 50 – 100 m; Lean et al. 2019) the derivation of morphology ancillaries should be revisited and evaluated. Using a representative canyon proportion (W/R) in MORUSES can strongly affect model behaviour and hence is important for NWP and climate modelling. However, for climate services, where it may be crucial to represent processes at street-canyon scale, accuracy of parameters linked to urban geometry (e.g. H/W) is required.

- Aerodynamic roughness: Current MORUSES model formulations effectively restrict the maximum allowed displacement height to the mean building height in the grid box. However, z_d can be large in cities with tall buildings and/or a large heterogeneity of building heights. Reflecting the impact of increased drag on flow profiles is crucial for the modelling of heat exchange between surface and atmosphere, which affects the estimation of thermal stress/comfort in urban environments. For applications in high-rise cities, it needs to be explored whether methods that account for building-height variations are more suitable.
- Urban vegetation modelling: Even small vegetation fractions in cities can strongly affect the energy partitioning (Loridan and Grimmond 2012a; 2012b), and accurate information on LAI and its seasonal variation is crucial to model evaporation correctly. Urban vegetation and water can be included by assuming a certain amount of both are present in the source urban data when it is processed into the JULES land-cover ancillaries. The urban vegetation and water are then represented although aggregated within the other (non-urban) JULES surface types in the grid box. Representing urban vegetation as separated from the USEB in JULES removes important process interactions, such as the impact of street trees on canyon aerodynamics (e.g. increased surface roughness; flow-regime changes), radiation balance (e.g. more shading; reduced sky-view factor; impact on bulk canyon albedo/emissivity) and heat-flux partitioning. These could be important in scenario studies, for example related to urban greening strategies as mitigation measures for thermal stress. Grimmond et al. (2011) found that ULSM performance of turbulent heat fluxes has a larger variability if vegetation is treated on separate tiles compared with using an integrated approach. The potential of further model improvements by including vegetation in MORUSES should be explored (Young et al. 2015; Young 2018).

The study documented the temporal (seasonal, diurnal) variability of model skill and the necessity to evaluate ULSMs over a range of meteorological conditions. Contrasting the effect of different resolution ancillaries on the

results showed that for integrated urban services that are informed by climate-model output an appropriate down-scaler should be used with suitable resolution land-cover information rather than using the output from the driving model directly. Hence, the configuration chosen needs to be appropriate for the purpose of the task. It also needs to be explored how the offline results will translate to coupled (online) simulations of UM–JULES over similar (multi-year) evaluation periods. There is also further need to evaluate the models in other high-density cities, in particular with tall-building canopies, and, once current modelling capabilities have been expanded, for different seasonality and types of anthropogenic forcing.

Acknowledgements We thank Martin Best at the Met Office as well as Jessica Brown, Allyn Treshansky and George Meachim at the University of Reading for their support for this work. We also acknowledge the expertise and assistance from the National Centre for Atmospheric Science Computational Modelling Services (NCAS/CMS).

Funding information This work and its contributors were supported by the UK–China Research & Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund (High Res City; AJYG-DX4P1V). Funding from the European Union’s Horizon 2020 Research & Innovation Programme (PRIMAVERA project; grant agreement no. 641727) was received (PLV), as is support for PCM through the University of Reading’s Research Endowment Trust Fund.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: JULES science configurations

The JULES–UKV (OS40) science configuration uses the same JULES settings as described by Bush et al. (2020) for RAL0. Differences in the atmospheric configuration and data assimilation used in OS40 and RAL0 do not affect the results of the offline test.

In addition to the urban schemes, there are other differences between the JULES–GL7.0 and JULES–UKV science configurations, in particular regarding the large-scale hydrology/runoff settings (cf. Section 6.4), the treatment of super-saturated soil layers and vertical gradients of soil suction. While JULES–GL7.0 uses a single albedo for plants for both the direct and diffuse beams,

JULES–UKV produces the albedos separately. Other differences concern some of the default parameters for inland water (‘lake’ tile), e.g. albedos and heat capacity. As the JULES–GL7.0 run (HAD–B) has only 2% lake in London (cf. Table 1a), the effect of this is negligible. As snow–fall rates were not measured at the sites (cf. Table 5, Appendix 4) and there were only few days with settled snow (Swindon only) during the evaluation period, differences of some of the snow parameters (e.g. snow-covered albedos for trees, fresh snow density) will not have an impact.

Appendix 2: MORUSES roughness parameters

Model formulations

MORUSES provides the option to compute the displacement height (z_d) and roughness length for momentum (z_0) internally using the Macdonald et al. (1998) method (Section 3.3). The displacement height is obtained from:

$$z_d = \left(1 - \frac{W}{R} \alpha_m^{(WR^{-1}-1)}\right) H \quad (8)$$

and the roughness length for momentum is:

$$z_0 = \left(1 - \frac{z_d}{H}\right) \exp \left[- \left(0.5 \beta_m \frac{C_D}{\kappa^2} \left(1 - \frac{z_d}{H}\right) \times \frac{2}{\pi} \frac{H}{W} \frac{W}{R}\right)^{-\frac{1}{2}} \right] H, \quad (9)$$

where $\alpha_m = 4.43$ and $\beta_m = 1.0$ are fit coefficients derived for staggered arrays (Macdonald et al. 1998), $C_D = 1.2$ is the corresponding drag coefficient, $\kappa = 0.4$ is the von Kármán constant, and the expression $2\pi^{-1}HW^{-1}WR^{-1}$ (Eq. 9) is equal to the frontal-area fraction, λ_f (cf. Eq. 3). MORUSES sets the minimum roughness length for momentum to that of the material roughness length ($z_{0m} = 0.05$ m). Both z_0 and z_d take the same value for the two urban tiles.

The roughness length for heat (z_h) is computed separately for the canyon and roof tiles using their respective bulk resistances, r_{bulk} (Porson et al. 2010a), resulting in:

$$z_h = \left[(z_{T,q} + z_0) \left(H \exp \left(\frac{\kappa^2 r_{\text{bulk}}}{\ln(z_{u,v} z_0^{-1} + 1)} \right) \right)^{-1} \right] H, \quad (10)$$

where $z_{T,q}$ and $z_{u,v}$ are the model forcing heights (Section 4). MORUSES imposes an overall lower limit of $z_h = 10^{-30}$ m. Additionally, for the isolated-flow regime

($H/W \leq 1/3$), there is an upper limit of the material roughness length for heat $z_h = z_{0h} = 0.1z_{0m}$.

Sensitivity study

Roughness parameters computed by MORUSES in response to changing W/R and H/W (Section 6.3; Fig. 10) are shown in Fig. 12. As the building height ($H = 19.81$ m) remains fixed, $z_d = 16.0$ m is constant at constant W/R for varying H/W (Fig. 12b; cf. Eq. 8), while changing with canyon fraction for $H/W = \text{const.}$ (Fig. 12a). In this test scenario, for low canyon fractions ($W/R \leq 0.25$), $z_0 = z_{0m}$ before increasing as z_d decreases.

The response of z_h to varying morphology parameters is shown individually for the canyon and roof tiles and as their weighted average (\bar{z}_h). In agreement with the variability of the corresponding grid box Q_H (Fig. 10b, e), the variation of \bar{z}_h is much larger in response to variations of canyon fraction ($0.01 \leq W/R \leq 0.99$) compared with the canyon aspect ratio ($0.15 \leq H/W \leq 3$); four orders of magnitude compared with one (Fig. 12a, b). In both tests, the response of z_h and \bar{z}_h to parameter changes is non-monotonic, which is partially triggered by the modelling constraints MORUSES puts on z_0 and z_h . Over the parameter ranges explored, the ratio of \bar{z}_h/z_0 remains larger than the value of 10^{-7} used in the Best-1T scheme (Section 3.3). However, with increasing canyon fraction \bar{z}_h/z_0 decreases and approaches the Best-1T value (Fig. 12a).

Appendix 3: Initial soil states

Soil characteristics reached after the spin-up period for CTRL-M/B, UKV-M and HAD-B at London/KCL and Swindon (cf. Section 6) in terms of soil temperature and soil wetness in the four soil layers are shown in Fig. S8 (Online Resource). Soil wetness is represented as the mass of soil water (liquid or solid) as a fraction of the water content at saturation.

The CTRL runs in London (region T) have notably higher soil temperatures in all layers compared with the test configurations and correspondingly are slightly drier than UKV-M and HAD-B (Fig. S8a). The differences in the initial soil wetness in London, however, are much smaller than those in Swindon (Fig. S8b). Here, HAD-B with the highest vegetation fraction has notably wetter soil below depths of 0.25 m and lower soil temperatures.

Appendix 4: Observed variables and instrumentation

Observed JULES forcing and evaluation variables (cf. Section 4) and instrumentation at the London/KCL and Swindon sites are given in Table 5 and described more fully in Ward et al. (2013), Kotthaus and Grimmond (2014a, b) and (Bjorkegren et al. 2015).

Fig. 12 MORUSES roughness length for momentum, z_0 , and displacement height, z_d (top) and roughness length for heat, z_h (bottom) for (a) varying W/R (H/W fixed) and (b) varying H/W (W/R fixed). The canyon- z_h value cut off in (a) for $W/R = 0.01$ is 10^{-30} m. \bar{z}_h is the weighted average of the canyon and roof z_h . The ratio $\bar{z}_h/z_0 = 10^{-7}$ (dashed line; bottom plots) prescribed in Best 1T, and mean building height, H (dashed line; top plots), are indicated

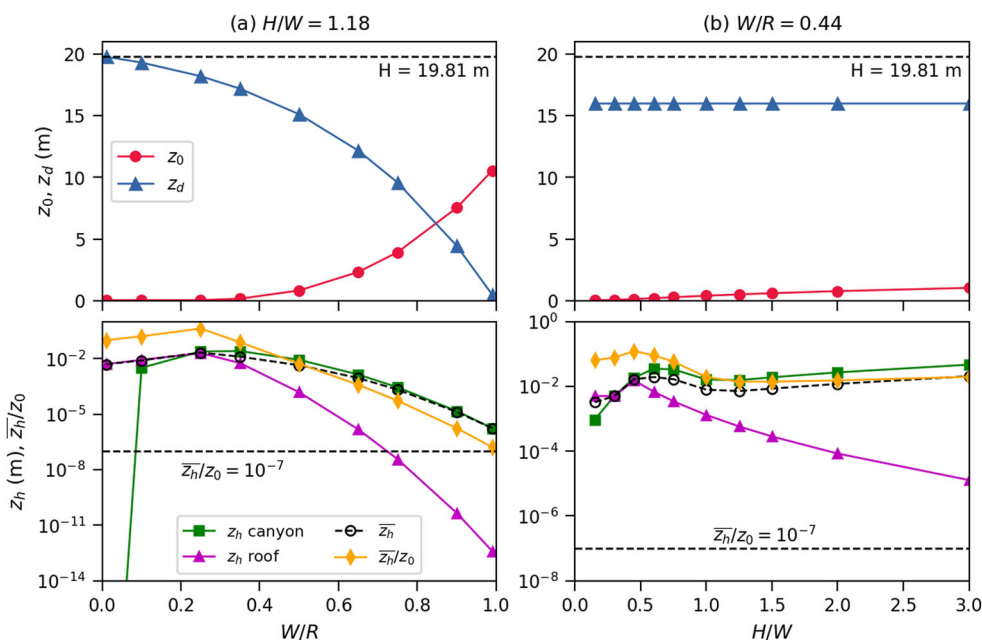


Table 5 Overview of (a) forcing and (b) evaluation variables observed in London ('L'; Kotthaus and Grimmond 2014a; Kotthaus and Grimmond 2014b; Bjorkegren et al. 2015) and Swindon ('S'; Ward et al. 2013). *SW*, shortwave; *LW*, longwave

Variable		Units	Instrument
(a) Forcing data			
Incoming SW radiation	$K \downarrow$	W m^{-2}	^L Kipp and Zonen CNR1/CNR4 ^S Hukseflux Thermal Sensors NR1
Incoming LW radiation	$L \downarrow$	W m^{-2}	^L Kipp and Zonen CNR1/CNR4 ^S Hukseflux Thermal Sensors NR1
Rainfall rate	RR	$\text{kg m}^{-2} \text{ s}^{-1}$	^L Vaisala WXT510/520; CSI raingauge ^S Casella CEL tipping bucket raingauge
Snowfall rate	SR	$\text{kg m}^{-2} \text{ s}^{-1}$	^{L,S} not observed
Air temperature	T_{air}	K	^{L,S} Vaisala WXT510/520
Wind speed	U_h	m s^{-1}	^{L,S} Vaisala WXT510/520
Specific humidity	q	kg kg^{-1}	^{L,S} Vaisala WXT510/520
Atmospheric pressure	P	Pa	^{L,S} Vaisala WXT510/520
(b) Evaluation data			
Outgoing SW radiation	$K \uparrow$	W m^{-2}	^L Kipp and Zonen CNR1/CNR4 ^S Hukseflux Thermal Sensors NR1
Outgoing LW radiation	$L \uparrow$	W m^{-2}	^L Kipp and Zonen CNR1/CNR4 ^S Hukseflux Thermal Sensors NR1
Net all-wave radiation	Q^*	W m^{-2}	^L Kipp and Zonen CNR1/CNR4 ^S Hukseflux Thermal Sensors NR1
Turbulent sensible heat flux	Q_H	W m^{-2}	^L CSI CSAT3; ^S Gill R3
Turbulent latent heat flux	Q_E	W m^{-2}	^L CSI CSAT3; ^S Gill R3; ^{L,S} Licor 7500

References











- Ao X, Grimmond CSB, Ward HC, Gabey AM, Tan J, Yang XQ, Liu D, Zhi X, Liu H, Zhang N (2018) Evaluation of the Surface Urban Energy and Water Balance Scheme (SUEWS) at a dense urban site in Shanghai: Sensitivity to anthropogenic heat and irrigation. *J Hydrometeorol* 19(12):1983–2005. <https://doi.org/10.1175/JHM-D-18-0057.1>
- Baklanov A, Grimmond CSB, Carlson D, Terblanche D, Tang X, Bouchet V, Lee B, Langendijk G, Kolli R, Hovsepyan A (2018) From urban meteorology, climate and environment research to integrated city services. *Urban Clim* 23:330–341. <https://doi.org/10.1016/j.uclim.2017.05.004>
- Barlow J, Best M, Bohnenstengel SI, Clark P, Grimmond S, Lean H et al (2017) Developing a research strategy to better understand, observe and simulate urban atmospheric processes at kilometre to sub-kilometre scales. *Bull Am Meteorol Soc* 98:ES261–ES264. <https://doi.org/10.1175/BAMS-D-17-0106.1>
- Best MJ (2005) Representing urban areas within operational numerical weather prediction models. *Bound-Lay Meteorol* 114(1):91–109. <https://doi.org/10.1007/s10546-004-4834-5>
- Best MJ (2006a) Progress towards better weather forecasts for city dwellers: from short range to climate change. *Theor Appl Climatol* 84:47–55. <https://doi.org/10.1007/s00704-005-0143-2>
- Best MJ, Grimmond CSB (2016) Modeling the partitioning of turbulent fluxes at urban sites with varying vegetation cover. *J Hydrometeorol* 17(10):2537–2553. <https://doi.org/10.1175/JHM-D-15-0126.1>
- Best MJ, Grimmond CSB, Villani MG (2006) Evaluation of the urban tile in mooses using surface energy balance observations. *Bound-Lay Meteorol* 118(3):503–525. <https://doi.org/10.1007/s10546-005-9025-5>
- Best MJ, Pryor M, Clark DB, Rooney GG, Essery RLH, Menard CB, Edwards JM, Hendry MA, Porson A, Gedney N, Mercado LM, Sitch S, Blyth E, Boucher O, Cox PM, Grimmond CSB, Harding RJ (2011) The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geosci Model Dev* 4(3):677–699. <https://doi.org/10.5194/gmd-4-677-2011>
- Bjorkegren A, Grimmond C, Kotthaus S, Malamud B (2015) CO₂ Emission estimation in the urban environment: Measurement of the CO₂ storage term. *Atmos Environ* 122:775–790. <https://doi.org/10.1016/j.atmosenv.2015.10.012>
- Bohnenstengel SI, Hendry MA (2016) Report on implementation and evaluation of MORUSES in the UKV (PS37). Met Office, Key Deliverable Report pp 1–52
- Bohnenstengel SI, Evans S, Clark PA, Belcher S (2011) Simulations of the London urban heat island. *Q J Roy Meteor Soc* 137(659):1625–1640. <https://doi.org/10.1002/qj.855>
- Bush M, Allen T, Bain C, Boutle I, Edwards J, Finnenkoetter A, Franklin C, Hanley K, Lean H, Lock A, Manners J, Mittermaier M, Morcrette C, North R, Petch J, Short C, Vosper S, Walters D, Webster S, Weeks M, Wilkinson J, Wood N, Zerroukat M (2020) The first Met Office Unified model–JULES Regional Atmosphere and Land configuration, RAL1. *Geosci Model Dev* 13(4):1999–2029. <https://doi.org/10.5194/gmd-13-1999-2020>
- CEH (1990). Land Cover Map of Great Britain 1990. Centre for ecology & hydrology, United Kingdom, Online resource, <https://www.ceh.ac.uk/services/land-cover-map-1990>
- Clark DB, Mercado LM, Sitch S, Jones CD, Gedney N, Best MJ, Pryor M, Rooney GG, Essery RLH, Blyth E, Boucher O, Harding RJ, Huntingford C, Cox PM (2011) The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics. *Geosci Model Dev* 4(3):701–722. <https://doi.org/10.5194/gmd-4-701-2011>

- Cox PM (2001) Description of the TRIFFID dynamic global vegetation model. Hadley Centre, Met Office, Bracknell, UK, Technical note 24
- Dou J, Grimmond S, Cheng Z, Miao S, Feng D, Liao M (2019) Summertime surface energy balance fluxes at two Beijing sites. *Int J Climatol* 39(5):2793–2810. <https://doi.org/10.1002/joc.5989>
- DUKES (2003) Digest of United Kingdom Energy Statistics 2003. Department of trade and industry's energy group, United Kingdom, DTI report
- Esch T, Schenk A, Ullmann T, Thiel M, Roth A, Dech S (2011) Characterization of land cover types in TerraSAR-x images by combined analysis of speckle statistics and intensity information. *IEEE T Geosci Remote Sens* 49(6):1911–1925. <https://doi.org/10.1109/TGRS.2010.2091644>
- Esch T, Heldens W, Hirner A, Keil M, Marconcini M, Roth A, Zeidler J, Dech S, Strano E (2017) Breaking new ground in mapping human settlements from space—The Global Urban Footprint. *ISPRS J Photogram Remote Sens* 134:30–42. <https://doi.org/10.1016/j.isprsjprs.2017.10.012>
- Esch T, Bachofer F, Heldens W, Hirner A, Marconcini M, Palacios-Lopez D, Roth A, Üreyen S, Zeidler J, Dech S, Gorelick N (2018) Where we live—a summary of the achievements and planned evolution of the global urban footprint. *Remote Sens* 10(6)
- Essery R, Best M, Cox P (2001) MOSES 2.2 Technical documentation. Met Office Hadley Centre, Hadley Centre Technical Note 30
- Evans S, Hudson-Smith A, Batty M (2006) 3-D GIS: Virtual London and beyond. *Cybergeo Europ J Geogr* 359. <https://doi.org/10.4000/cybergeo.2871>
- Gabey AM, Grimmond CSB, Capel-Timms I (2019) Anthropogenic heat flux: advisable spatial resolutions when input data are scarce. *Theor Appl Climatol* 135(1):791–807. <https://doi.org/10.1007/s00704-018-2367-y>
- Gilham R, Hendry M, Edwards JM, Rooney GG, Folberth GA (2019). Coupling the JULES Land Surface Model to the Unified Model. Met Office, United Kingdom, Unified Model Documentation Paper C03
- Grimmond CSB (1992) The suburban energy balance: Methodological considerations and results for a mid-latitude west coast city under winter and spring conditions. *Int J Climatol* 12(5):481–497. <https://doi.org/10.1002/joc.3370120506>
- Grimmond CSB, Best M, Barlow J, Arnfield AJ, Baik JJ, Baklanov A, Belcher S, Bruse M, Calmet I, Chen F, Clark P, Dandou A, Ereli E, Fortuniak K, Hamdi R, Kanda M, Kawai T, Kondo H, Krayenhoff S, Lee SH, Limor SB, Martilli A, Masson V, Miao S, Mills G, Moriawaki R, Oleson K, Porson A, Sievers U, Tombrou M, Voogt J, Williamson T (2009) Urban surface energy balance models: Model characteristics and methodology for a comparison study. Springer, Berlin, pp 97–123. https://doi.org/10.1007/978-3-642-00298-4_11
- Grimmond CSB, Blackett M, Best MJ, Barlow J, Baik JJ, 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 SH, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Pigeon G, Porson A, Ryu YH, Salamanca F, Shashua-Bar L, Steeneveld GJ, Tombrou M, Voogt J, Young D, Zhang N (2010) The international urban energy balance models comparison project: First results from phase 1. *J Appl Meteor Climatol* 49(6):1268–1292. <https://doi.org/10.1175/2010JAMC2354.1>
- Grimmond CSB, Blackett M, Best MJ, Baik JJ, Belcher SE, Beringer J, Bohnenstengel SI, Calmet I, Chen F, Coultts A, Dandou A, Fortuniak K, Gouvea ML, Hamdi R, Hendry M, Kanda M, Kawai T, Kawamoto Y, Kondo H, Krayenhoff ES, Lee SH, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Ooka R, Pigeon G, Porson A, Ryu YH, Salamanca F, Steeneveld G, 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(2):244–272. <https://doi.org/10.1002/joc.2227>
- Harman IN, Barlow JF, Belcher SE (2004) Scalar fluxes from urban street canyons part II: model. *Bound-Lay Meteorol* 113(3):387–410. <https://doi.org/10.1007/s10546-004-6205-7>
- Holm E, Forbes R, Lang S, Magnusson L, Malardel S (2016) New model cycle brings higher resolution. *European Centre for Medium-Range Weather Forecasts, ECMWF newsletter* 147, 14–19
- Iamarino M, Beevers S, Grimmond CSB (2012) High-resolution (space, time) anthropogenic heat emissions: London 1970–2025. *Int J Climatol* 32(11):1754–1767. <https://doi.org/10.1002/joc.2390>
- Kanda M, Inagaki A, Miyamoto T, Gryschka M, Raasch S (2013) A new aerodynamic parametrization for real urban surfaces. *Bound-Lay Meteorol* 148(2):357–377. <https://doi.org/10.1007/s10546-013-9818-x>
- Kent CW, Grimmond S, Barlow J, Gatey D, Kotthaus S, Lindberg F, Halios CH (2017) Evaluation of urban local-scale aerodynamic parameters: Implications for the vertical profile of wind speed and for source areas. *Bound-Lay Meteorol* 164(2):183–213. <https://doi.org/10.1007/s10546-017-0248-z>
- Kent CW, Lee K, Ward HC, Hong JW, Hong J, Gatey D, Grimmond S (2018) Aerodynamic roughness variation with vegetation: analysis in a suburban neighbourhood and a city park. *Urb Ecosys* 21(2):227–243. <https://doi.org/10.1007/s11252-017-0710-1>
- Kent CW, Grimmond CSB, Gatey D, Barlow JF (2018a) Assessing methods to extrapolate the vertical wind-speed profile from surface observations in a city centre during strong winds. *J Wind Eng Ind Aerodyn* 173:100–111. <https://doi.org/10.1016/j.jweia.2017.09.007>
- Kotthaus S, Grimmond CSB (2014a) Energy exchange in a dense urban environment – Part I: Temporal variability of long-term observations in central London. *Urban Clim* 10:261–280. <https://doi.org/10.1016/j.uclim.2013.10.002>
- Kotthaus S, Grimmond CSB (2014b) Energy exchange in a dense urban environment – Part II: Impact of spatial heterogeneity of the surface. *Urban Clim* 10:281–307. <https://doi.org/10.1016/j.uclim.2013.10.001>
- Lean HW, Barlow JF, Halios CH (2019) The impact of spin-up and resolution on the representation of a clear convective boundary layer over London in order 100-m grid-length versions of the Met Office Unified Model. *Q J Roy Meteor Soc* 145(721):1674–1689. <https://doi.org/10.1002/qj.3519>
- Lindberg F, Grimmond CSB, Gabey A, Huang B, Kent CW, Sun T, Theeuwes NE, Järvi L, Ward HC, Capel-Timms I, Chang Y, Jonsson P, Krave N, Liu D, Meyer D, Olofson KFG, Tan J, Wästberg D, Xue L, Zhang Z (2018) Urban Multi-scale Environmental Predictor (UMEP): An integrated tool for city-based climate services. *Environ Modell Softw* 99:70–87. <https://doi.org/10.1016/j.envsoft.2017.09.020>
- Loridan T, Grimmond CSB (2012a) Multi-site evaluation of an urban land-surface model: intra-urban heterogeneity, seasonality and parameter complexity requirements. *Q J Roy Meteor Soc* 138(665):1094–1113. <https://doi.org/10.1002/qj.963>
- Loridan T, Grimmond CSB (2012b) Characterization of energy flux partitioning in urban environments: Links with surface seasonal properties. *J Appl Meteor Climatol* 51(2):219–241. <https://doi.org/10.1175/JAMC-D-11-038.1>
- Macdonald R, Griffiths R, Hall D (1998) An improved method for the estimation of surface roughness of obstacle arrays. *Atmos Environ* 32(11):1857–1864. [https://doi.org/10.1016/S1352-2310\(97\)00403-2](https://doi.org/10.1016/S1352-2310(97)00403-2)
- Martilli A, Clappier A, Rotach MW (2002) An urban surface exchange parameterisation for mesoscale models. *Bound-Lay Meteorol* 104(2):261–304. <https://doi.org/10.1023/A:1016099921195>

- Masson V (2000) A physically-based scheme for the urban energy budget in atmospheric models. *Bound-Lay Meteorol* 94(3):357–397. <https://doi.org/10.1023/A:1002463829265>
- Oke TR, Mills G, Christen a, Voogt JA (2017) *Urban Climates*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/9781139016476>
- Porson A, Clark PA, Harman IN, Best MJ, Belcher SE (2010a) Implementation of a new urban energy budget scheme in the MetUM. Part I: Description and idealized simulations. *Q J Roy Meteor Soc* 136(651):1514–1529. <https://doi.org/10.1002/qj.668>
- Porson A, Clark PA, Harman IN, Best MJ, Belcher SE (2010b) Implementation of a new urban energy budget scheme into MetUM. Part II: Validation against observations and model intercomparison. *Q J Roy Meteor Soc* 136(651):1530–1542. <https://doi.org/10.1002/qj.572>
- Sailor DJ (2011) A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment. *Int J Climatol* 31(2):189–199. <https://doi.org/10.1002/joc.2106>
- Sailor DJ, Vasireddy C (2006) Correcting aggregate energy consumption data to account for variability in local weather. *Environ Modell Softw* 21(5):733–738. <https://doi.org/10.1016/j.envsoft.2005.08.001>
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res-Atmos* 106(D7):7183–7192. <https://doi.org/10.1029/2000JD900719>
- United Nations (2018) *World Urbanization Prospects: The 2018 Revision*. United Nations, Department of Economic and Social Affairs, Online edition
- Vidale P, Roberts M, Baker A, Bellouin N, Muller O, Schiemann R, Thornhill G, Vanniere B (in. prep.) Description of the resolution hierarchy of HadGEM3-GC3.1 model as used in the HighResMIP atmosphere-only experiments. In preparation for Geosci Model Dev
- Walters D, Baran AJ, Boutle I, Brooks M, Earnshaw P, Edwards J, Furtado K, Hill P, Lock A, Manners J, Morcrette C, Mulcahy J, Sanchez C, Smith C, Stratton R, Tennant W, Tomassini L, Van Weverberg K, Vosper S, Willett M, Browse J, Bushell A, Carslaw K, Dalvi M, Essery R, Gedney N, Hardiman S, Johnson B, Johnson C, Jones A, Jones C, Mann G, Milton S, Rumbold H, Sellar A, Ujiie M, Whitall M, Williams K, Zerroukat M (2019) The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geosci Model Dev* 12(5):1909–1963. <https://doi.org/10.5194/gmd-12-1909-2019>
- Ward H, Kotthaus S, Järvi L, Grimmond CSB (2016) Surface Urban Energy and Water Balance Scheme (SUEWS): Development and evaluation at two UK sites. *Urban Clim* 18:1–32. <https://doi.org/10.1016/j.uclim.2016.05.001>
- Ward HC, Evans JG, Grimmond CSB (2013) Multi-season eddy covariance observations of energy, water and carbon fluxes over a suburban area in Swindon, UK. *Atmos Chem Phys* 13(9):4645–4666. <https://doi.org/10.5194/acp-13-4645-2013>
- Warren E, Charlton-Perez C, Kotthaus S, Lean H, Ballard S, Hopkin E, Grimmond S (2018) Evaluation of forward-modelled attenuated backscatter using an urban ceilometer network in London under clear-sky conditions. *Atmos Environ* 191:532–547. <https://doi.org/10.1016/j.atmosenv.2018.04.045>
- Wiltshire AJ, Duran Rojas C, Edwards J, Gedney N, Harper AB, Hartley A, Hendry M, Robertson E, Smout-Day K (2020) JULES–GL7: The Global Land Configuration of the Joint UK Land Environment Simulation version 7.0 and 7.2. *Geosci Model Dev* 13(2):483–505. <https://doi.org/10.5194/gmd-13-483-2020>
- WMO (2018) Guide for urban integrated hydro-meteorological, climate and environmental services. Part 1a: Concept and methodology. World Meteorological Organization The GAW Urban Research Meteorology and Environment (GURME) project. Online document
- Young DT (2018) Representing urban vegetation in weather and climate models (under revision). PhD thesis, University of Reading
- Young DT, Clark P, Hendry M, Barlow J (2015) Modelling radiative exchange in a vegetated urban street canyon model. In: *Proceedings of ICUC9 9th International Conference on Urban Climate*, Toulouse, France

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Denise Hertwig¹  · Sue Grimmond¹  · Margaret A. Hendry²  · Beth Saunders¹  · Zhengda Wang^{1,3}  · Marine Jeoffrion^{1,4} · Pier Luigi Vidale¹  · Patrick C. McGuire^{1,5,6}  · Sylvia I. Bohnenstengel⁷  · Helen C. Ward⁸  · Simone Kotthaus⁹ 

¹ Department of Meteorology, University of Reading, Reading, UK

² Met Office, Exeter, UK

³ Present address: Fengxian District Weather Office, Shanghai Meteorological Service, Shanghai, China

⁴ Present address: Météo-France, Toulouse, France

⁵ Department of Geography, Environmental Science, University of Reading, Reading, UK

⁶ National Centre for Atmospheric Science, Reading, UK

⁷ Met Office@Reading, Meteorology Building, University of Reading, Reading, UK

⁸ Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria

⁹ Institut Pierre Simon Laplace (IPSL), École Polytechnique, Palaiseau, France